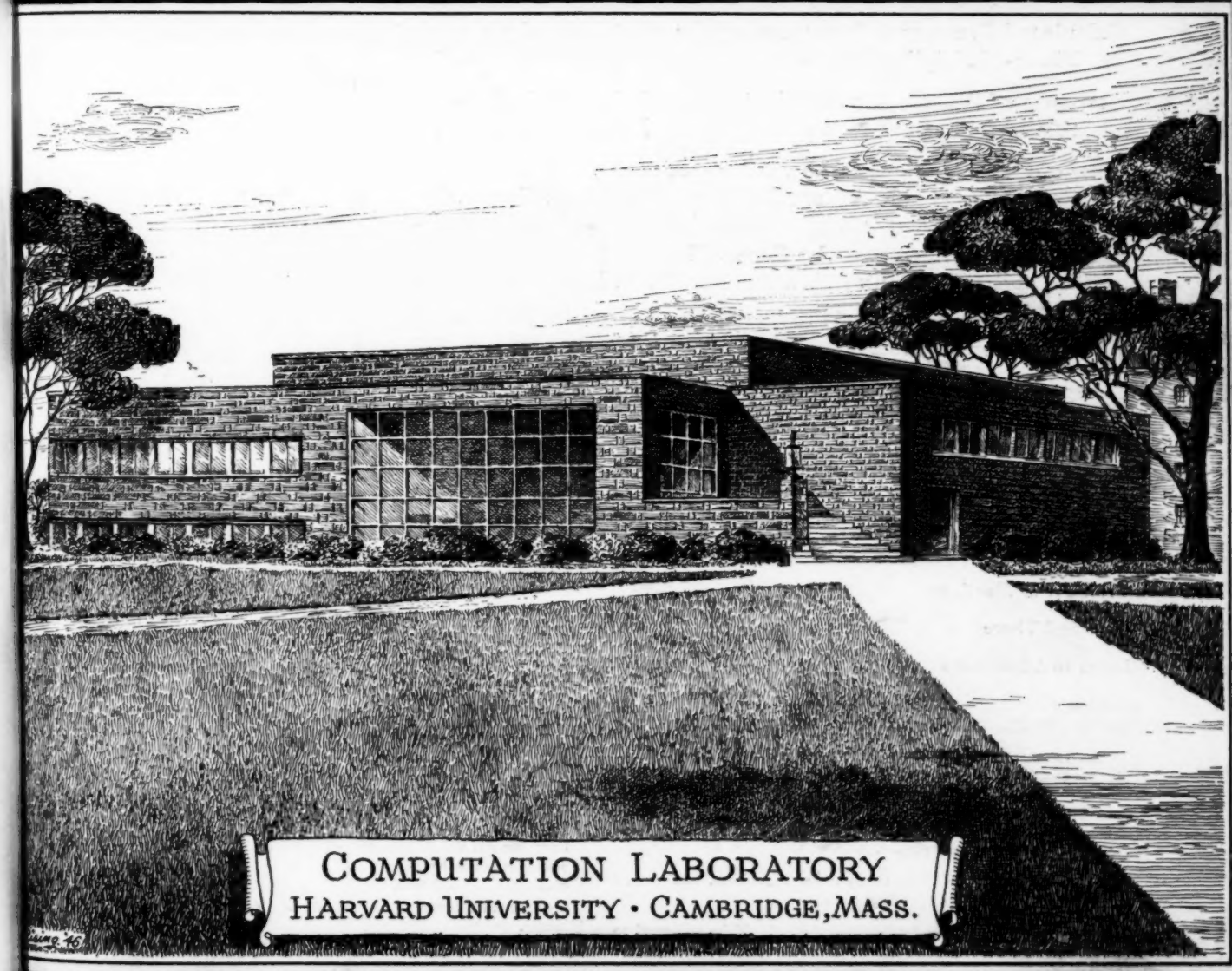
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 17, No. 10 OCTOBER, 1946

In This Issue

Supplementary Bibliography of Electron Microscopy MARY E. RATHBUN, MARY J. EASTWOOD, AND ORLAN M. ARNOLD 759

Contributed Original Research

Electronic Spectroscopy GEORGE C. SZIKLAI AND ALFRED C. SCHROEDER 763

Calculated Frequency Spectrum of the Shot Noise from a Photo-Multiplier Tube R. D. SARD 768

Response of Biased, Saturated Linear and Quadratic Rectifiers to Random Noise DAVID MIDDLETON 778

Flow of Gases in Pipes at Low Pressures GORDON P. BROWN, ALBERT DINARDO, GEORGE K. CHENG, AND THOMAS K. SHERWOOD 802

Absorption of Microwaves by Gases. II JOHN E. WALTER AND W. D. HERSHBERGER 814

Electrical Observations of the Austenite-Martensite Transformation in Steel ANDREW W. MCREYNOLDS 823

Energy Build-Up in Magnetrons LLOYD P. HUNTER 833

Measured Impedance of Cylindrical Dipoles D. D. KING 844

Elementary Treatment of Longitudinal Debunching in a Velocity Modulation System EUGENE FEENBERG 852

In Every Issue

Calendar of Meetings 801

Here and There 856

Index to Advertisers xxx

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
S. H. CALDWELL
CLARK GOODMAN
SELIG HECHT
W. JAMES LYONS
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND

Publication of the American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*
MARGARET S. GRIFFIN, *Publications Manager*
THEODORE VORBURGER, *Business Manager*
JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

PAUL E. KLOPSTEG, *Chairman*
WALLACE WATERFALL, *Secretary*
GEORGE B. PEGRAM, *Treasurer*
R. B. DOW
L. A. DUBRIDGE
W. F. FAIR, JR.
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
W. H. MARKWOOD, JR.
A. H. PFUND
JOHN T. TATE
L. W. TAYLOR
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case School of Applied Science, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members of scientific societies	$5.00	$5.70
To all others	7.00	7.70

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 16, 1945—$101.50
Yearly back number rate: **$7.70**
Single copies: **$0.70** each.

Changes of address and proofs should be addressed to the Publications Manager.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1946, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 17, Number 10 October, 1946

Supplementary Bibliography of Electron Microscopy

BY MARY E. RATHBUN, MARY J. EASTWOOD, AND ORLAN M. ARNOLD
Engineering Division, Chrysler Corporation, Detroit, Michigan

IN an extensive survey of the literature on the subject of electron microscopy a number of publications have been discovered which were not included in the excellent bibliographies of C. Marton and S. Sass.[1] In view of the request expressed by a number of people interested in electron microscopy to present a list of these articles which have technical points of interest bearing on technique, preparation of samples, instrumentation, and results, the authors have prepared this supplementary bibliography in two alphabetized sections. The first group includes publications which have information on various phases of electron microscopy. The second section gives the patent literature that has appeared to date. The presentation of the latter section has been prompted by the expressed opinion that many working in this field of research would like to become familiar with different stages of development of the electron microscope as reflected in both foreign and domestic patents. The domestic patents have been checked with The Official Gazette of the United States Patent Office and the foreign patents with the information in the Chemical Abstracts. For the convenience of the reader, translations have been made for the titles of publications and patents appearing in foreign languages.

The authors wish to acknowledge the aid of Eleanor V. Wright, Librarian,* in helping to locate a number of the publications reviewed for this bibliography.

ELECTRON MICROSCOPE REFERENCES NOT INCLUDED IN BIBLIOGRAPHY OF C. MARTON AND S. SASS

Ahearn, A. J. and Becker, J. A. Electron microscope studies of thoriated tungsten. *Phys. Rev.* **54**, 448–458 (1938).

Barnes, R. B. and Burton, C. J. Electron microscope. *Am. Dyestuff Rep.* **31**, 254–262 (1942); correction **31**, 313–314 (1942).

Barrett, C. S. Seeing metals at high magnification. *Mach. Design* **15**, 138–140 (November, 1943).

Benjamin, M. Field emission microscope. *J. Soc. Glass Tech.* **24**, 93–96 (1940); abstracted in *Chem. Ind. Eng.* **59**, 314–315 (1940).

Bensen, I. B. Electron microscopy. *Gen. Elec. Rev.* **47**, No. 12, 6–14 (1944).

Boersch, H. The shadow-microscope, a new electron-supermicroscope. *Naturwiss.* **27**, 418 (1939).

Brüche, E. Electron microscope. *Naturwiss.* **20**, 49 (1932).

Brüche, E. Electron optics and electron microscopy. *AEG Mitt.* No. 2, 45–47 (February, 1934).

Brüche, E. Ten years of development. *AEG-Forsch. Jahrb., Sonderheft, Ubermikroskop* **7**, 2–8 (1940).

Brüche, E. The two-pole system as aim of pure electrical image apparatus. *AEG-Forsch. Jahrb., Sonderheft, Ubermikroskop* **7**, 9–14 (1940).

Brüche, E. and Gölz, E. Introduction of the object and the

[1] C. Marton and S. Sass, J. App. Phys. **14**, 522–31 (1943); **15**, 575–9 (1944); **16**, 373–8 (1945).

* Engineering Division, Chrysler Corporation, Detroit, Michigan.

plate. *AEG-Forsch. Jahrb., Sonderheft, Ubermikroskop* **7**, 60–66 (1940).

Brüche, E. and Johannson, H. Electron optics and electron microscope. *Naturwiss.* **20**, 353–358 (1932).

Burgers, W. G. Metallographic investigation with the electron microscope. *Polytech. Weekblad.* **33**, 17–18, 38–40 (1939).

Dautz, D. Supermicroscope and paint research. *Farbe u. Lack, Centralblatt* 402–403 (1938).

Dorgelo, H. B. Introduction to electron microscopes of different types. *Nederlandsch Tijdschrift voor Natuurkunde* **7**, 157–170 (1940).

Gaede. The electron microscope. *Deut. Med. Wochschr.* **66**, 858–860 (1940).

Hendricks, S. B., Wildman, S. G., and McMurdie, H. F. Morphology of latex particles as shown by electron micrographs. *India Rubber World* **110**, 297–300 (1944).

Henneberg, W. The supermicroscope with electrostatic lenses. *Elekrotech. Zeits.* **61**, 773–776 (1940).

Hillier, J. The electron microscope in the determination of particle size characteristics. *Am. Soc. Test. Mat.*, Symposium on New Methods for Particle Size Det'n. in Subsieve Range 90–94 (1941).

Hillier, J. Electron microscopy. *Can. Chem. Process Ind.* **28**, 728–736 (1944).

Hock, C. W. and McMurdie, H. F. Structure of the wool fiber as revealed by the electron microscope. *Am. Dyestuff Rep.* **32**, 433–436, 451–454 (1944).

Hoffmann, K. Characteristics of optical illustrations by means of electrons and applications (electron microscopy). *Kolloid Zeits.* **89**, 59–76 (1939).

Humbert, R. P. Particle shape and the behavior of clay as revealed by the electron microscope. *Bull. Am. Cer. Soc.* **21**, 260–263 (1942).

Jakus, M. A., Hall, C. E., and Schmitt, F. O. Electron microscope observations of clam muscle fibrils. *J. Am. Chem. Soc.* **66**, 313–314 (1944).

Jelley, E. E. Structure of the developed silver image as revealed by the electron microscope. *J. Phot. Soc. Am.* **8**, 283–289 (1942).

Kalden, H. Investigations with the supermicroscope. *Chem.-Zeit.* **64**, 129–133 (1940).

Kinder, E. The magnetic sectional lens supermicroscope and some applications in colloid chemistry. *Kolloid Zeits.* **95**, 326–36 (1941).

Knaysi, G. Width and origin of bacterial flagella; electron microscope studies. *Science* **95**, 406–407 (1942).

Knoll, M. and Theile, R. Electron scanning for forming the image of the structure of surfaces and thin layers. *Zeits. f. Physik* **113**, 260–280 (1939).

Kronig, R. Theoretical basis of electron optics. *Nederlandsch Tijdschrift voor Naturrkunde* **7**, 171–178 (1940).

Loofbourow, J. R. Borderland problems in biology and physics; electron microscopy. *Rev. Mod. Phys.* **12**, 331–333 (1940).

Mahl, H. Metallurgical, metal chemical and bacteriological investigations with the electrostic supermicroscope. *Chem. Fabrik* **14**, 279–280 (1941).

Mahl, H. Stereoscopic photographs with the electrostatic supermicroscope. *Naturwiss.* **28**, 264 (1940).

Malov, N. N. Electron microscope. *Uspekh. Fiz. Nauk.* **13**, 367–384 (1933).

Martin, L. C. The present position regarding the development of the electron microscope. *J. Soc. Glass Tech.* **24**, 97–100 (1940).

Morton, G. A. Survey of research accomplishments with the RCA electron microscope. *RCA Rev.* **6**, 131–166 (1941).

Morton, G. A. and Ramberg, E. G. Point projector electron microscope. *Phys. Rev.* **56**, 705 (1939).

Müller, E. W. Further observations with the field electron microscope. *Zeits. f. Physik* **108**, 668–680 (1938).

Pupko, S. L. Electron microscope. *Uspekh. Fiz. Nauk.* **24**, 487–513 (1940).

Recknagel, A. Defects of electron lenses. *Jahrb. AEG-Forsch., Sonderheft, Ubermikroskop* **7**, 15–22 (1940).

Rees, A. L. G. Electron microscope. *J. Soc. Chem. Ind.* **60**, 335–337 (1941).

Rees, A. L. G. Electron microscopy. *Sci. J. Roy. College of Science* **12**, 1–15 (1942).

Rhea, H. E. The RCA electron microscope. *Wallerstein Labs. Commun.* **4**, 99–106 (1941).

Rhea, H. E. New electron microscope. *Science* **93**, 357–358 (1941).

Ross, S. Orientation of micelles in soap fibers. *J. Phys. Chem.* **46**, 414–417 (1942).

Schmitt, F. O. X-ray and electron microscope studies of the structure of collagen fibers. *Am. Leather Chem. Assn. J.* **39**, 430–437 (1944); Discussion 437–441.

Semmler, E. The structure of aluminum oxide films used in surface structure investigations by the contact film method (electron microscopy). *Zeits. f. Metallkunde* **34**, 229–231 (1942)—Translation: Great Britain Air Ministry RTP_3 Translation No. 1880.

Smith, P. C. and Picard, R. G. Electron microscopes. *Radio News* **32**, 41–43 (1944) (November).

Stäger, A. Electronic and scanning microscopes. *Elec. Rev. (London)* **124**, 157–158 (1939).

Stanley, W. M. Viruses and the electron microscope. *Chronica Botan.* **7**, 291–294 (1943).

Triau, L. The most recent realization of the electron microscope. *Rev. gén. élec.* **44**, 841–847 (1938).

Wilson, W. The electron microscope. *Electronic Eng.* **16**, 414–420 (1944).

Wilson, W. Electron microscope. Part I: The Instrument. *Elec. Rev. (London)* **134**, 218–222 (1944). Part II: Specimen Mounting and Typical Results. *Elec. Rev. (London)* **134**, 254–257 (1944).

Zworykin, V. K. Electron microscopy in chemistry. *Electronics* **16**, 64–68 (1943).

Zworykin, V. K. Image formation by electrons. *Cold Spring Harbor Symposia Quant. Biol.* **9**, 194–196 (1941); Discussion 196–197.

Zworykin, V. K. Image formation by electrons. *Science in Progress* [3] 69–107 (1942).

Zworykin, V. K., Morton, G. A., Ramberg, E. G., Hillier, J., and Vance, A. W. *Electron optics and the electron microscope* (John Wiley & Sons, Inc., New York, 1945), 766 pp.

Development and uses of the electron microscope. *Chemische Fabrik* **11**, 478–479 (1938).

Electron microscopes. *Electronics* **11**, 30–33 (November, 1938).

Electron microscopes for production, research and analysis. *Electronics* **17**, 184 (June, 1944).

Electron microscope: the Siemens electron microscope. *Engineering* **146**, 474–475 (October, 1938).

A new electron microscope. *J. App. Phys.* **11**, 629–630 (1940).

Polystyrene aids electron microscope. *Mod. Plastics* **21**, 124–125, 160 (December, 1943).

Technique for the electron microscopic examination of encapsulated bacteria. *Paper Tr. J.* **113**, 33 (October, 1941).

Inside rubber. *Tech. Rev.* **46**, 479–481 (1944).

PATENTS

Ardenne, M. von. Magnetic objective for electron microscopes. U.S. 2,243,403. May 27, 1941.

Ardenne, M. von. Electron scanning microscope. U.S. 2,253,542. Aug. 26, 1941.

Ardenne, M. von. Ultraviolet or electronic microscope. U.S. 2,273,235. Feb. 17, 1942.

Ardenne, M. von. Arrangement for controlling the anode potential and the lens currents of magnetic electron microscopes. Ger. 718,988. March 5, 1942.

Ardenne, M. von. Electron microscope. Ger. 732,791. Feb. 11, 1943.

Ardenne, M. von. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,356,633. Aug. 22, 1944.

Ardenne, M. von. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,361,722. Oct. 31, 1944.

Ardenne, M. von and Borries, B. von. Electron scanning microscope. U.S. 2,241,432. May 13, 1941.

Boersch, H. (Assigned to General Electric Company) Electron microscope. U.S. 2,238,577. April 15, 1941.

Boersch, H. Electron tubes for electron-diffraction apparatus or electron microscopes. Ger. 725,847. Aug. 20, 1942.

Boersch, H. (Assigned to Allgemeine Elektricitäts-Gesellshaft) Electron microscope for taking pictures with light and dark films. Ger. 734,736. Mar. 25, 1943.

Boersch, H., Brüche, E., and Mahl, H. (Assigned to General Electric Company) Electron microscope. U.S. 2,259,994. Oct. 21, 1941.

Boersch, H. and Mahl, H. (Assigned to General Electric Company) Electron Microscope. U.S. 2,249,453. July 15, 1941.

Boersch, H. and Mahl, H. (Assigned to Allgemeine Elektricitäts-Gesellshaft) Highly magnifying electron microscope. Ger. 729,687. Nov. 26, 1942.

Boersch, H. and Mahl, H. (Assigned to Allgemeine Elektricitäts-Gesellshaft) Electron microscope with two or more single lenses. Ger. 735,873. April 22, 1943. Addn. to Ger. 733,345.

Borries, B. von. Electron microscope (Vested in the Alien Property Custodian). U.S. 2,301,987. Nov. 17, 1942.

Borries, B. von and Müller, H. O. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron optical device. U.S. 2,270,112. Jan. 13, 1942.

Borries, B. von, Müller, H. O., and Ruska, E. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,356,551. Aug. 22, 1944.

Borries, B. von and Ruska, E. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Magnetic lens for electron microscopes. U.S. 2,209,669. July 30, 1940.

Borries, B. von and Ruska, E. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,268,539. Dec. 30, 1941.

Borries, B. von and Ruska, E. Electron-optical lens (Vested in the Alien Property Custodian). U.S. 2,305,761. Dec. 22, 1942.

Borries, B. von and Ruska, H. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Method for examining objects in an air space by means of electron microscopes. U.S. 2,256,191. Sept. 16, 1941.

Borries, B. von, Schuchmann, H., and Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,266,717. Dec. 16, 1941.

Brüche, E. (Assigned to Allgemeine Elektricitäts-Gesellshaft) High-power electron microscope. Ger. 732,690. Feb. 11, 1943.

Brüche, E. (Assigned to Allgemeine Elektricitäts-Gesellshaft) Electron microscope. Ger. 734,090. March 11, 1943.

Brüche, E. and Gölz, E. (Assigned to Allgemeine Elektricitäts-Gesellshaft) Apparatus for introducing objects into an electron microscope. Ger. 716,058. Dec. 11, 1941.

Burkhardt, E. (Assigned to Allgemeine Elektricitäts-Gesellshaft) Adjusting arrangement for electron microscopes. Ger. 724,183. July 9, 1942.

Hillier, J. (Assigned to RCA Company) Electron microscope specimen chamber. U.S. 2,272,843. Feb. 10, 1942.

Hillier, J. (Assigned to RCA Company) Projection lens for an electron microscope. U.S. 2,323,328. July 6, 1943.

Hillier, J. (Assigned to RCA Company) Scanning microscope. U.S. 2,330,888. Oct. 5, 1943.

Hillier, J. (Assigned to RCA Company) Electron microscope. U.S. 2,354,263. July 25, 1944.

Knoll, M., Houtermans, F. G., and Schulze, W. Electron microscope. U.S. 2,131,536. Sept. 27, 1938.

Krause, F. Electron microscope vacuum system. U.S. 2,266,218. Dec. 16, 1941.

Krause, F. (Assigned to Gewerkschaft Mathias Stinnes) Device for introducing objects into an electron microscope. Ger. 719,448. March 12, 1942. Addn. to Ger. 708,778.

Krause, F. Preparation of specimens for examination with an electron microscope. Ger. 720,213. April 2, 1942.

Krause, F. Method for preparing specimens of cellular material for studying in an electron microscope. Ger. 720,949. April 23, 1942. Addn. to Ger. 720,213.

Krause, K. Device for introducing objects into an electron microscope. Ger. 708,778. June 19, 1941.

Krause, K. Means for mounting objects in electron microscopes. U.S. 2,264,210. Nov. 25, 1941.

Mahl, H., Pendzich, A., and Steudel, E. (Assigned to General Electric Company) Electron microscope. U.S. 2,260,041. Oct. 21, 1941.

Marton, L. (Assigned to RCA Company) Electron microscope. U.S. 2,220,973. Nov. 12, 1940.

Marton, L. (Assigned to RCA Company) Electronic and light microscope. U.S. 2,233,286. Feb. 25, 1941.

Marton, L. (Assigned to RCA Company) Shielded electronic microscope. U.S. 2,301,303. Nov. 10, 1942.

Marton, L. (Assigned to RCA Company) Electron microscope. U.S. 2,347,328. Apr. 25, 1944.

Ploke, M. (Assigned to Zeiss Ikon Aktiengesellschaft) Method of electron-microscopically investigating subjects. U.S. 2,219,113. Oct. 22, 1940.

Rajchman, J. A. (Assigned to RCA Company) Method of focusing electron microscopes. U.S. 2,348,031. May 2, 1944.

Rajchman, J. A. and Snyder, R. L. (Assigned to RCA Company) Electron optical instrument. U.S. 2,372,450. Mar. 27, 1945.

Ramberg, E. G. (Assigned to RCA Company) Art of electron microscopy. U.S. 2,347,965. May 2, 1944.

Ramberg, E. G. and Morton, G. A. (Assigned to RCA Company) Electron microscope. U.S. 2,271,990. Feb. 3, 1942.

Ramo, S. (Assigned to General Electric Company) Electron microscope. U.S. 2,281,325. Apr. 28, 1942.

Ramo, S. (Assigned to General Electric Company) Electron microscope. U.S. 2,363,359. Nov. 21, 1944.

Ruska, E. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Shielded electron microscope. U.S. 2,234,281. March 11, 1941.

Ruska, E. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,266,082. Dec. 16, 1941.

Ruska, E. (Assigned to "Fides Gesellschaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electronic microscope. U.S. 2,272,353. Feb. 10, 1942.

Ruska, E. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electronic microscope. U.S. 2,274,215. Feb. 24, 1942.

Ruska, E. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,301,975. Nov. 17, 1942.

Ruska, E. (Assigned to Siemens & Halske Aktiengesellshaft-Gesellshaft) Electron microscope. Ger. 734,791. March 25, 1943.

Ruska, E., Hentschel, E., and Schuchmann, H. Electronic microscope (Vested in the Alien Property Custodian) U.S. 2,305,458. Dec. 15, 1942.

Ruska, E. and Ruska, H. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U. S. 2,267,137. Dec. 23, 1941.

Ruska, E. and Schuchmann, H. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electronic microscope. U.S. 2,277,024. March 17, 1942.

Ruska, E., Schuchmann, H., Ruska, H., and Müller, H. O. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,370,373. Feb. 27, 1945.

Ruska, H. (Assigned to Siemens & Halske Aktiengesellshaft-Gesellshaft) Mount films for electron microscope. Ger. 709,733. July 17, 1941.

Ruska, H. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Method of focusing electron microscopes. U.S. 2,256,171. Sept. 16, 1941.

Schuchmann, H. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscope. U.S. 2,215,979. Sept. 24, 1940.

Schuchmann, H., Borries B. von, and Ruska, H. Electronic microscope (Vested in the Alien Property Custodian). U.S. 2,305,459. Dec. 15, 1942.

Schuchmann, H. and Neubert, U. (Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electronic microscope. U.S. 2,247,524. July 1, 1941.

Snyder, R. L., Jr. (Assigned to RCA Company) Scanning type of electron microscope. U.S. 2,330,930. Oct. 5, 1943.

Snyder, R. L., Jr. (Assigned to RCA Company) Scanning type of electron microscope. U.S. 2,348,030. May 2, 1944.

Sukumlyn, T. W. Electron microscope system. U.S. 2,243,362. May 27, 1941.

Young, C. J. (Assigned to RCA Company) Electron microscope object chamber. U.S. 2,347,348. Apr. 25, 1944.

Young, C. J. (Assigned to RCA Company) Electron microscope viewing chamber. U.S. 2,356,963. Aug. 29, 1944.

No Authors Given

(Assigned to "Fides Gesellshaft für die Verwaltung und Verwertung von Gewerblichen Schutzrechten m. b. H.") Electron microscopes. Brit. 517,704. Feb. 6, 1940.

(Assigned to Allgemeine Elektricitäts-Gesellshaft) Electron-optic method for investigating mixtures of crystals. Ger. 713,740. Oct. 16, 1941.

(Assigned to Verwaltungs-Gesellshaft m. b. H.) Electron ultramicroscope. Fr. 849,648. Nov. 28, 1940.

Contributed Original Research

Electronic Spectroscopy

GEORGE C. SZIKLAI AND ALFRED C. SCHROEDER
RCA Laboratories, Princeton, New Jersey
(Received April 12, 1946)

According to Einstein, the electrons leave a photoelectric surface with an initial velocity proportional to the frequency of the impinging radiation. Consequently, when a sawtooth potential is applied to a photoelectric cell, illuminated by a monochromatic light, the current flow starts at a threshold potential corresponding to the wave-length of the light, and then gradually increases more or less linearly toward a saturation value. When light of assorted wave-lengths is applied, the total current is the summation of the currents for the individual wave-lengths, each of which currents has its threshold at a different potential. By double differentiation an indication of the spectral distribution can be observed directly on an oscilloscope which is deflected synchronously with the potential applied to the photoelectric cell. The accuracy of the indication is limited by the indefiniteness of the threshold values and non-linearity of the currents above the threshold potentials in available photoelectric cells. With S4 commercial surfaces, filters of the three primary colors can be easily differentiated, but with improved surfaces considerably better resolutions may be obtained and further improvement may be had by lowering the temperature of the photoelectric surface. The method lends itself readily to color matching, since two similar devices supplying their signal with opposite polarity will give zero output when the two colors match, while it provides a direct indication of the color and magnitude of mismatch when a signal is obtained. The output signal can be utilized for relays or other control devices in industrial processes.

INTRODUCTION

FIRST we may define spectroscopy, according to Sir Arthur Schuster,[1] as an investigation of radiations with respect to the wave-lengths of their constituent parts.

Spectroscopy, as a branch of physics, came into prominence after the spectroscope and its applications were introduced by Kirchoff and Bunsen in 1859, almost two centuries after Newton discovered that a prism breaks white light into its components. The optical spectroscope of Kirchoff and Bunsen was improved in many ways for greater convenience (such as the Brown, Amici, and Kessler direct line spectroscopes with compound prisms and by the use of gratings) and since 1919 through the addition of photo-cells for quantitative recording.[2] However, these instruments were still using purely optical principles, namely, refraction and diffraction, for the analysis of the radiation.

The present scheme uses electronic properties for the analysis and it borrows its principle from the Einstein's law of photoelectric emission, and thus we call the instrument electronic.

THE PRINCIPLE OF OPERATION

The photoelectric equation of Einstein[3] postulates that the initial kinetic energy of a photoelectron is a linear function of the frequency of the liberating radiation according to the relation of

$$mv^2/2 = h\nu - \phi,$$

where the left side represents the kinetic energy, h is Planck's constant, ν is the frequency of the radiation, and ϕ is the work function. The kinetic energy is also equal to the product of the electronic charge and the accelerating voltage; thus we may write

$$eV = h\nu - \phi,$$

where e is the charge and V is the accelerating potential required to impart to an electron the velocity V when the electron started in vacuum from rest. In the last equation all the quantities except V and ν are constant; thus we may see

[1] A lecture delivered at Royal Institution in 1882.
[2] K. S. Gibson *et al*, Nat. Sci. Pap. Bur. Stand. No. 349, 1919.
[3] Ann. d. Physik **17**, 132 (1905).

that the potential at which electron flow starts or the threshold potential is a direct function of the frequency of the radiation of a given surface.

According to the Fermi-Dirac distribution of the energy among the conduction electrons, as

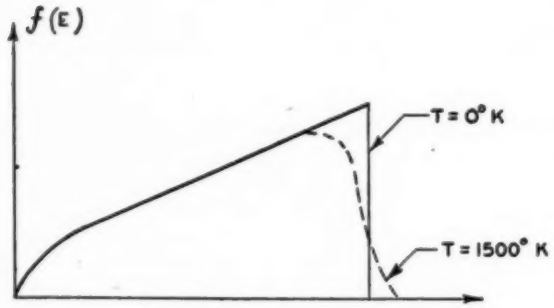


FIG. 1. The Fermi-Dirac energy distribution curve.

shown in Fig. 1, the number of electrons per unit energy is zero above a certain energy level which is lower than the work function of the metal. Figure 1 shows the distribution at 0°K and at 1500°K where the rounding off, approaching Maxwellian distribution, is caused by thermionic effect. The curve at room temperature is between the two curves shown.

Applying this information to the current in a photo-cell which is connected to a variable retarding potential, we find that if certain wavelength radiation is applied on the surface and the retarding potential is reduced to the threshold potential current flow begins. As the retarding potential is reduced further, the current increases almost linearly until saturation occurs. Figure 2 shows the effect of the radiation of two discreet bands on the current in a photo-cell as the stopping or retarding potential is reduced. The first dotted line of curve *B* shows the current produced by the higher frequency radiation, while the dotted line on the right corresponds to the lower frequency. The solid line indicates the current *vs.* stopping potential relation when both radiations are present. The curvature at the foot of each of the curves is caused by thermionic emission, and in some cases by non-definite work function or threshold. Such is the case with compound or non-uniform surfaces.

If curve *B* is differentiated, with respect to the stopping potential, we obtain curve *C*. A second differentiation yields curve *A*, again reproducing the original spectra.

THE CIRCUIT

Figure 3 shows the simplest form of an apparatus performing the above described functions. A periodically variable potential (preferably sawtooth shaped) is supplied to a suitable photoelectric cell (1) by a source (2). The center potential is established by a battery or other d.c. source (3). The variation of the current changes the potential across a resistor (4) in the circuit. A differentiating circuit consisting of a capacitor (5) and a resistor (6) supplies the differentiated signal to the grid of a vacuum tube amplifier (7). This amplifier is followed by a similar differentiating circuit-feeding amplifier (8). The amplified twice-differentiated signal is then applied to the vertical deflection system of a cathode-ray oscilloscope (9) which is deflected horizontally with the same periodicity as the potential across the cell (1) is varied.

Figure 4 shows a simple circuit for differentiation with two photoelectric cells (10) and (11) connected in series and biased with the aid of slightly different d.c. sources (12) and (13) at different center potentials. The periodical wave is applied by source (2′) through a center tapped

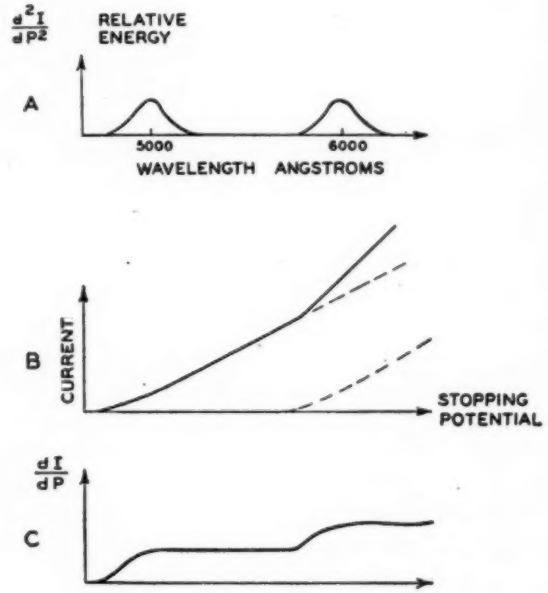


FIG. 2. Graphical representation of the photo-current differentiation.

transformer (14), and the differentiated output is obtained across resistor (4′). This output then is differentiated again by capacitor (5′) and resistor (6′), amplified by amplifier (8′) and applied to the oscilloscope (9′) similarly to the way shown in Fig. 1.

Figure 5 shows the manner in which differentiation is taking place by the circuit shown in Fig. 4. Curve *A* shows the voltage developed by one photoelectric cell when it is exposed to a light source of red and blue content as shown by curve *D*. Curve *B* is the same as *A* with opposite polarity, except it is delayed by the above-mentioned bias difference. The resultant potential (the sum of the two potentials) that appears across resistor (4′) is shown as curve *C*, which is substantially the differential of curve *A*. If signal *C* is differentiated the resultant potential will be similar to the curve *D*.

A special circuit is shown in Fig. 6, which is similar to Fig. 4, except that one of the cells observes the light to be tested, while the other

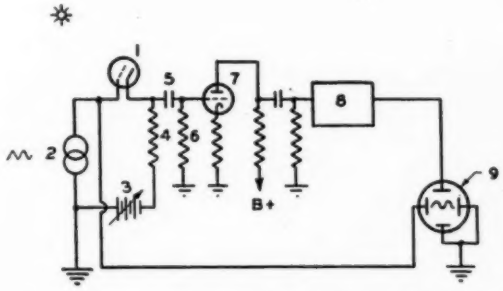


FIG. 3. Simple circuit with double R–C differentiation.

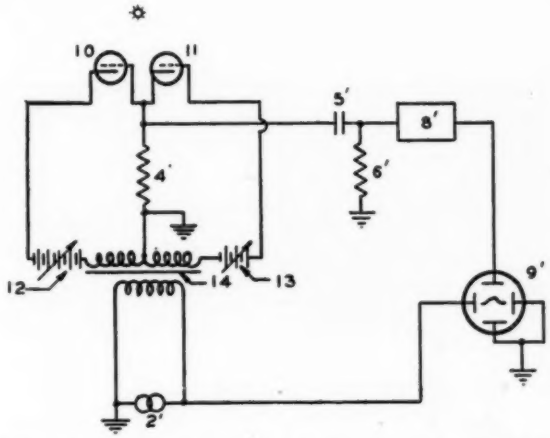


FIG. 4. Circuit with differentiation in push-pull cells.

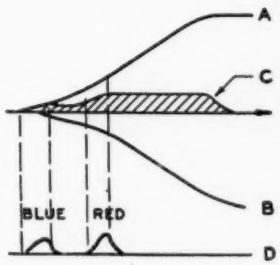


FIG. 5. Differentiation by push-pull action.

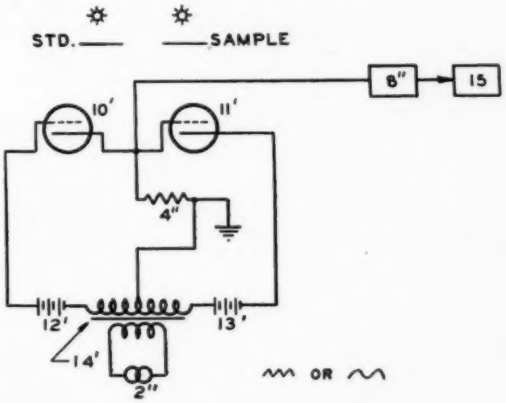


FIG. 6. Color matching diagram.

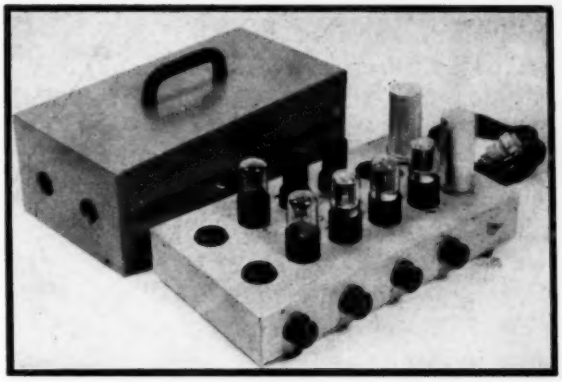


FIG. 7. The experimental unit.

provides a signal from a standard to be matched. The two bias potentials (12′) and (13′) are equal, and thus if the two light sources (or reflectors) have the same spectral distribution, the currents through cells (10′) and (11′) are the same, and the output signal across resistor (4″) is zero. The amplifier (8″) provides an actuating signal to relay 15 only if the two colors are not matched. This circuit finds a particular application in

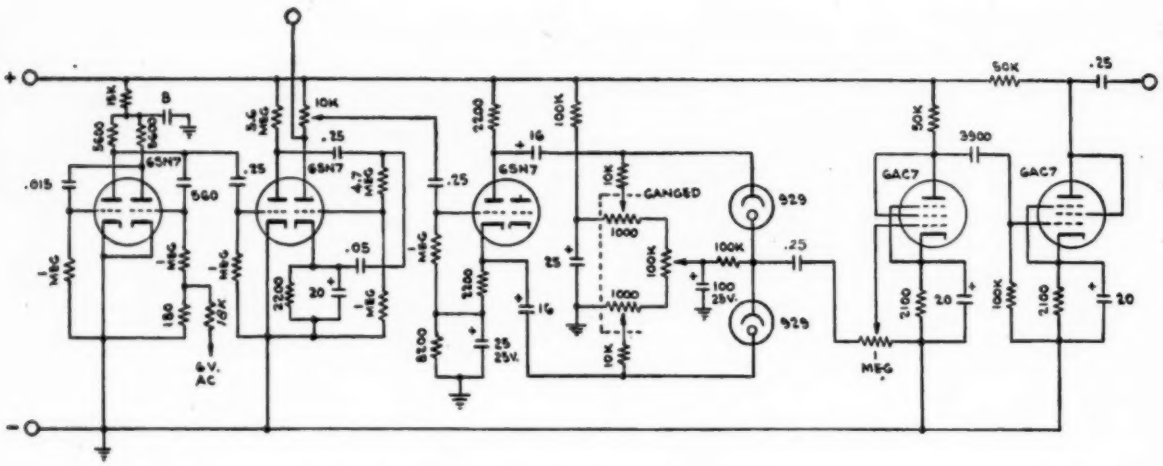


FIG. 8. Circuit of the experimental unit.

continuous dyeing processes (such as dyeing of yarns, filter sheets, etc.), where the output of the bath may be continuously matched and the equipment is stopped immediately if the material is mismatched with respect to the standard.

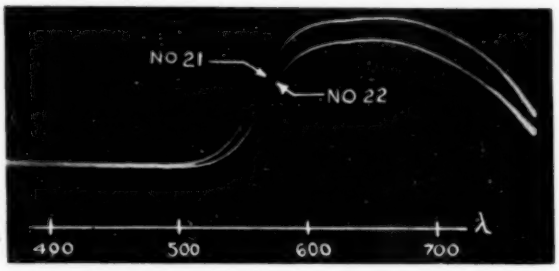
EXPERIMENTAL RESULTS

A simple set-up, similar to the one shown in Fig. 3, was made with a 919 photo-tube. The multiple threshold of the C_s-C_sO-Ag surface masked the effect to great extent to counteract the advantage of the great sensitivity of the surface. The 919 tube was replaced with a 929 type of photo-tube containing an S4 surface. This set-up was capable of differentiating the standard tristimulus colors. The results obtained were encouraging enough to construct a color analyzing assembly using two 929 tubes as shown in Fig. 7. The circuit diagram of the analyzer is shown in Fig. 8.

The color analyzer was checked with a monochromator and a shift of 150 angstrom units was easily seen on the face of the oscilloscope. This may not be considered the resolution of the equipment since the half-value points on the curves obtained from monochromatic light were approximately 600 angstrom units apart, but the color-matcher may be expected to differentiate between two sources different with 150 angstrom units. Figure 9 shows a double exposed oscillogram of the traces obtained from a Wratten No. 21 and a Wratten No. 22 filter. The saturation close to 7000 angstrom units clearly shows, and the slope of the curves provides a good indication of the resolution available with S4 surfaces. For the purpose of comparison, the density curves of the same filters are shown in Fig. 10.[4]

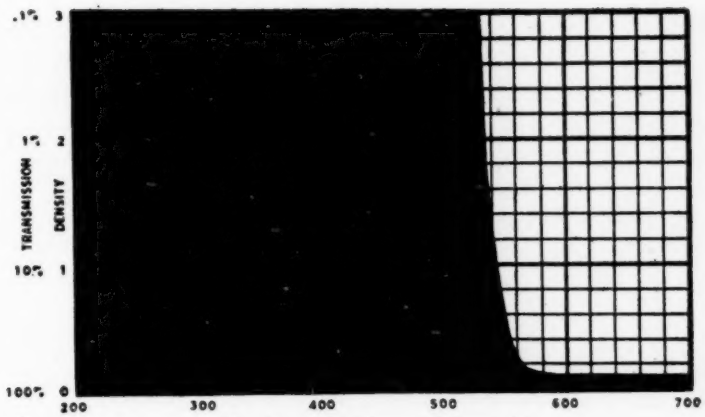
Considerable improvements may be made in the equipment if some of the limitations in the presently used commercial cells were removed. Because of the non-parallel construction of the photo-tubes, the collection of the electrons are not uniform through the stopping voltage range. A photoelectric cell made with flat parallel surfaces would eliminate the electron optical limitations, since electrons from the whole surface would be uniformly collected or retarded.

An obvious improvement may be obtained by lowering the temperature to reduce the thermionic effect and thus sharpen the threshold.

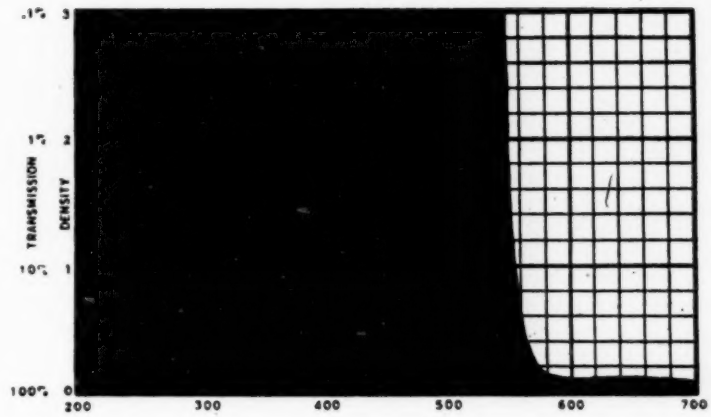


[FIG. 9. Transmission curves of Wratten No. 21 and 22 filters.

[4] *Wratten Light Filters* (Eastman Kodak Company, New York), sixteenth edition, p. 34.



No. 21. Monobromofluoresceine



No. 22. E2

FIG. 10. Published density curves of Wratten No. 21 and 22 filters.

Another step to obtain a sharp threshold is the use of pure metal surfaces eliminating the non-uniformity of the work function over the surface.

Frohlich[5] has obtained extremely sharp thresholds from thin alkali metal films presumably due to eliminating the effects of absorption of light and of electrons within the film.

[5] Ann. d. Physik **7**, 103 (1930).

While the present results are adequate for indications in industrial applications, further refinements may provide an instrument with greater precision.

The simplicity of the instrument and its convenient use due to its instantaneous application ought to be an incentive for further work in making the required refinements.

Calculated Frequency Spectrum of the Shot Noise from a Photo-Multiplier Tube*

R. D. SARD
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received April 4, 1946)

The mean square noise current per unit frequency range as a function of frequency, or "power spectrum," of the shot noise produced by a secondary-emission multiplier tube is calculated. A general expression for it, (3), is set up and then applied to the RCA 931 family of photo-multiplier tubes. These tubes are of interest as relatively intense sources of random noise of constant intensity up to high frequencies. When an electron enters the collecting space, it initiates a pulse of current in the anode circuit. In the 931 operated with 100 volts per stage, the pulse has a total duration of about 6×10^{-10} sec. and a half-width of about 1.5×10^{-10} sec. The electrons descended from the same photon do not enter the collecting space simultaneously but are spread out over a time of the order of 6×10^{-10} sec. Because of the non-zero duration of the pulses and the spread in their time of occurrence, the noise spectrum is essentially flat only up to several hundred Mc, falling rapidly at higher frequencies. The paucity of data and the approximations made in the calculation make our curve of noise output *versus* frequency significant only as regards order of magnitude and general trend. It does seem safe to predict that the noise intensity ought to be constant from zero up to about 100 Mc, begin to fall off appreciably between 100 Mc and 1000 Mc, and become very weak at higher frequencies. Partial confirmation is provided by unpublished measurements of Jastram at the Radio Research Laboratory, Harvard University, which show that at 30 Mc the noise output is still about the same as in the range 0 to 5 Mc. There is also some experimental indication that the output is about the same at 100 Mc and that it is much smaller at higher frequencies.

1. INTRODUCTION AND SUMMARY

THE subject of this paper is the frequency spectrum of the noise current in a vacuum secondary-emission multiplier tube. Under normal operating conditions, the fluctuations in collector potential associated with the noise fluctuations are small compared with the applied constant accelerating potential differences, and it is therefore legitimate to consider the noise current as independent of the external load. After the mean square noise current per unit frequency interval has been calculated on this assumption, the mean square noise voltage can be obtained from it by multiplying with the square of the absolute value of the load impedance and integrating over all frequencies.

We derive first a very general expression (Eq. (3)) for the mean square noise current per unit frequency interval for any type of multiplier tube, and then evaluate it for the RCA 931 family of photo-multiplier tubes. Approximations are made in the evaluation that are of a crudity commensurate with the meagerness of the available data on the properties of the secondary-emitting surfaces and the electron trajectories in these tubes, and with the variations in behavior from tube to tube. For this reason, the numerical results are significant for frequencies above several hundred megacycles only as regards order of magnitude and general trend. The calculation may, nevertheless, be of interest to physicists and radio engineers, not only because of the physical ideas involved, but also because of the possibility of using a tube like the 931 as a laboratory noise source for the study of receiver sensitivity. Because of its high gain, as explained below, this tube gives a large mean square noise current for a given average value of the anode current. And the noise is essentially white, i.e. its frequency spectrum has a constant intensity, from zero to at least one hundred megacycles.

Shot noise in a tube of any kind is caused by the fact that charge is brought to the collector in discrete lumps whose times of arrival are not correlated with each other. In a saturated diode, the individual electrons moving from cathode to anode constitute the lumps. Each such transit electron moves independently of the others, producing its independent transient current in the plate circuit. The piling-up of these transients produces the plate current. Because the transients occur at random, the plate current undergoes random fluctuations about its average value. In a secondary-emission multiplier tube,

* This paper is based on work done for the Office of Scientific Research and Development under Contract No. OEMsr-411 with the President and Fellows of Harvard College.

the phenomenon of secondary electron emission is used to magnify the primary current. By the use of a number of secondary-emitting stages in cascade, one primary electron is made to produce a large number of electrons in the last stage of multiplication. Thus, in the 931, which has nine stages of multiplication, one photon incident on the photo-sensitive cathode liberates a photoelectron which then produces hundreds of thousands of electrons in the ninth generation. These electrons are collected by the anode at almost the same instant. The independent lumps in which charge reaches the anode are, therefore, hundreds of thousands of times larger than in the case of a saturated diode. The effect is equivalent to an increase of the electronic charge by this factor in the formula for the shot noise of a saturated diode.

As an electron approaches the anode it induces a pulse of current in the external circuit to the anode. We shall refer to a current pulse due to an individual electron as an *elementary pulse*. The sum of the elementary pulses due to all the electrons descended from the same primary will be called a *burst*. As the primary photoelectrons occur at random, the bursts that they produce also occur at random. The bursts are the *random elements*, or lumps, which pile up randomly to form the anode current.

It is seen that in general the theory of shot noise falls naturally into two parts: a physical part concerned with the amplitude and shape in time of the random elements (in our problem, the bursts) and a statistical part concerned with the various mean values resulting from the random superposition of these elements.

We have to calculate the noise *power spectrum*, the mean square noise current per unit frequency interval as a function of frequency. The statistical part of the problem is disposed of by the well-known theorem[1] that the average power spectrum of the sum of a number of random elements is the sum of the power spectra of the individual elements (by "average" is meant the average over an ensemble of independent intervals, in each of which a Fourier analysis is made; the result is approached asymptotically by the spectrum of a single interval as its duration increases). We need only to calculate, then, the power spectra of the individual bursts. While considerable variations in the amplitudes of the bursts may be expected (and are, indeed, taken into account in our calculation), their shapes as functions of time should be nearly the same. We therefore calculate the power spectra of the bursts on the assumption that the bursts all have one typical shape.

The shape of a burst is determined by the shapes of the constituent elementary pulses and by their distribution in time. If the spread in time of occurrence of the elementary pulses is small compared to their duration, the elementary pulses simply pile up right on top of one another, and the burst shape is determined by the elementary pulse shapes. If the spread in time of occurrence is large compared to the duration of the elementary pulses (but still small enough to give considerable overlapping of the elementary pulses), the shapes of the elementary pulses are washed out, and the burst has the shape of the distribution function of time of occurrence of the elementary pulses. If the durations of the elementary pulses and the spread in their times of occurrence are of the same order of magnitude, both must be taken into account. It turns out that in the 931 a typical elementary pulse has a total duration of about 6×10^{-10} sec. and a duration between half-amplitude points of about 1.5×10^{-10} sec. The distribution curve for time of occurrence of the elementary pulses has a width between half-amplitude points of about 6×10^{-10} sec. Hence, both the duration of the elementary pulses and their spread in time of occurrence must be taken into account.

The non-zero duration of the elementary pulses is caused by the finite velocities with which the electrons approach the anode. As an electron approaches the anode, it induces on it an increasing amount of positive charge, which is supplied by a current in the external circuit. This current is equal to the rate of change of the charge induced on the anode, being a pulse of duration equal to the time interval during which an appreciable fraction of the electron's lines of force end on the anode. It is our elementary pulse. To calculate its shape, we need to determine: (1), the position and velocity of the electron as a function of the time; and (2), the

[1] Rayleigh, Phil. Mag. **28**, 460 (1889).

amount of charge induced on the anode by the electron as a function of its position. The first problem is the dynamical one of the motion of an electron in a given field of force; the second is an electrostatic one which can be solved in certain cases by the method of images.[2]

The spread in time of occurrence can be owing to: (1) fluctuating time lags in secondary emission, (2) spread in initial velocities of the secondary electrons. The spread in emission velocities has two results. The secondaries liberated by a particular impinging electron have different times of flight to the next electrode. Secondly, they reach it at different points, so producing a spatial spreading of the cascade. This spatial spreading entails variations in times of flight for different branches of the cascade. Reasons are given in Section III for believing that in the 931 the main cause of spread in time of occurrence of the elementary pulses is the spread in emission velocities as it affects the time of flight from a particular point of one electrode to the next electrode. The other effects just mentioned should involve smaller time spreads, and should, therefore, only affect the noise spectrum at higher frequencies.

The time elapsed between the arrival of the primary electron at the first multiplication electrode and the beginning of an elementary pulse is the sum of eight independently fluctuating inter-stage transit times. The most probable total elapsed time is therefore the sum of the most probable inter-stage transit times; and the mean square deviation of the total elapsed time is the sum of the mean square deviations of the inter-stage transit times. As eight is a large number from the point of view of the central limits theorem of statistics, the distribution curve of time of occurrence of the elementary pulses will be Gaussian.

[2] This paragraph appears to contradict the atomicity of electricity, as it implies that the induced charge varies continuously from zero to one electron during the flight of the transit electron. Professor Van Vleck has pointed out to the author that the paradox disappears when one considers the atomic nature of the external circuit. When we speak of the induced charge on a body, we are really referring to its electric polarization; the induced charge is the result of the displacement of the metallic conduction electrons. This polarization is a continuous variable. The quantum-mechanical justification for the use of electrostatic image theory when the inducing charge is a single electron has been given by J. Bardeen (Phys. Rev. **58**, 727 (1940)).

In our calculation of the burst spectrum, we assume that the elementary pulses constituting a burst all have essentially the same shape. We then use a well-known theorem in Fourier analysis[3] which, applied here, states that the spectrum of the burst is the product of the spectrum of the elementary pulse and the spectrum of the distribution curve of time of occurrence of the elementary pulses. This saves us the labor of computing the burst shape itself. We need only calculate the mean elementary pulse and the distribution curve of time of occurrence, find the Fourier transform of each, and then take their products.

The weakest point in our calculation for the 931 would seem to lie in the calculation of the spread in inter-stage transit time. No data were available on the secondary emission velocities from the surfaces actually used in these tubes, and very little was known about the electron trajectories. In order to get some idea of the orders of magnitude, we have used Malter's published data[4] on normal emission velocities from a silver-magnesium surface similar to that used in the 931's, applying them to the case of plane parallel electrodes. The result of the use of this crude approximation and of other less dubious ones, and of the physical variations from tube to tube, should make our results significant only as to order of magnitude and general trend. The initial falling off in noise output with rising frequency should, however, be given fairly accurately; and it does seem safe to predict that the noise spectrum should be essentially flat up to about 100 Mc and start to fall off appreciably in the range of 100 to 1000 Mc. Partial confirmation is provided by unpublished measurements of Jastram at the Radio Research Laboratory, Harvard University, which show that at 30 Mc the noise output is still about the same as in the range 0 to 5 Mc. There is also some experimental indication that the output is about the same at 100 Mc and that it is much smaller at higher frequencies.

[3] E.g., G. A. Campbell and R. M. Foster, *Fourier Integrals for Practical Applications* (Bell Telephone System Publications, New York, 1931), Monograph B-584, formula 202.

[4] L. Malter, Proc. I.R.E. **29**, 587 (1941).

II. GENERAL THEORY OF THE SHOT NOISE IN A MULTIPLIER TUBE

Let the elementary pulse due to the ith electron have a shape represented by the function $\Delta_i(t-t_i)$, where t_i is the time at which the pulse ends (electron reaches anode). If we normalize Δ_i to unit area, the elementary pulse is $e\Delta_i(t-t_i)$ where e is the electronic charge. The total anode current is

$$i(t)=\sum_i e\Delta_i(t-t_i). \tag{1}$$

This is the current to the anode through the external circuit, which includes internal capacitances.

We can group together the elementary pulses belonging to the same burst (i.e., descended from the same primary), and write

$$i(t)=\sum_k eM_kB_k(t). \tag{2}$$

The index k refers to the burst. M_k is the number of electrons in the burst; $B_k(t)$ is the shape of the kth burst, which we have normalized to unit area.

Expanding the current in a Fourier series of fundamental period T (long enough so that the number of harmonics in a small frequency range $\Delta\nu$ is proportional to $\Delta\nu$), we obtain

$$\langle P(\nu)\rangle_{Av}=2e\langle i(t)\rangle_{Av}\frac{\langle M^2\rangle_{Av}}{\langle M\rangle_{Av}}\cdot\langle|T\mathbf{B}(\nu)|^2\rangle_{Av}, \tag{3}$$

where $\mathbf{B}(\nu)$ is the Fourier transform of the burst shape $B(t)$:

$$\mathbf{B}(\nu)=\frac{1}{T}\int_0^T B(t)\exp(-i2\pi\nu t)dt, \tag{3a}$$

and $P(\nu)$ is the mean square noise current per unit frequency range. The derivation involves a verification of Rayleigh's result.[1] No physical assumption about the bursts is involved other than that they occur at random; they need not be similar in shape. Formulas (3) and (3a) are the basis for the calculations of this paper.

It is interesting to note from (3) that fluctuations in the multiplication factor increase the noise power[5] because of the factor $(\langle M^2\rangle_{Av}/\langle M\rangle_{Av})$. The effective size of the lumps of charge is not $e\langle M\rangle_{Av}$ but rather $(e\langle M^2\rangle_{Av}/\langle M\rangle_{Av})$.

For frequencies so low that the period is large compared to the burst duration, $B_k(t)$ can be replaced by a δ function, and

$$\mathbf{B}_k(\nu)=\frac{1}{T}\exp(-i2\pi\nu t_k); \tag{4}$$

where t_k is the time of occurrence of the burst. (3) becomes

$$\langle P(\nu)\rangle_{Av}=2e\langle i(t)\rangle_{Av}\frac{\langle M^2\rangle_{Av}}{\langle M\rangle_{Av}}. \tag{4a}$$

This special case of (3) has been derived by Shockley and Pierce.[6] For higher frequencies, the shape of the burst becomes important in determining $\mathbf{B}_k(\nu)$. Finally, when the frequency becomes so high that a large number of oscillations take place in a time in which $B_k(t)$ changes but little, $\mathbf{B}_k(\nu)$ becomes small, and approaches zero. In general, then, the noise spectrum is flat at zero frequency and approaches zero at high frequency (inversely as the square of the frequency), with an intermediate region in which its behavior is determined by the burst shape.

In practice the fluctuations in the potential of the anode relative to the preceding stage are small compared to the applied d.c. potential differences. This justifies our calculation of a given noise current independent of the load. The measured noise *voltage* will depend on the impedance of the load, which, as already remarked, includes the coupling of the anode to the other electrodes and leads. Calling the total impedance $Z(\nu)$, we have for the mean square voltage across the load

$$\langle v^2\rangle_{Av}=2T\int_{\nu=0}^{\infty}|Z(\nu)|^2\cdot\langle|I(\nu)|^2\rangle_{Av}\cdot d\nu, \tag{5}$$

where $I(\nu)$ is the transform of $i(t)$. This gives for

[5] Ziegler (Physica **3**, 307 (1936)) has pointed out how measurement of the mean square noise current from a one-stage multiplier gives the mean square number of secondary electrons liberated by one primary. It may be worth remarking that measurement of the higher moments of the output noise would give the higher moments of the distribution function for the number of secondaries per primary. In this way, the probabilities of emission of 0, 1, 2, 3, ··· secondaries per primary could be determined.

[6] W. S. Shockley and J. R. Pierce, Proc. I.R.E. **26**, 321 (1938).

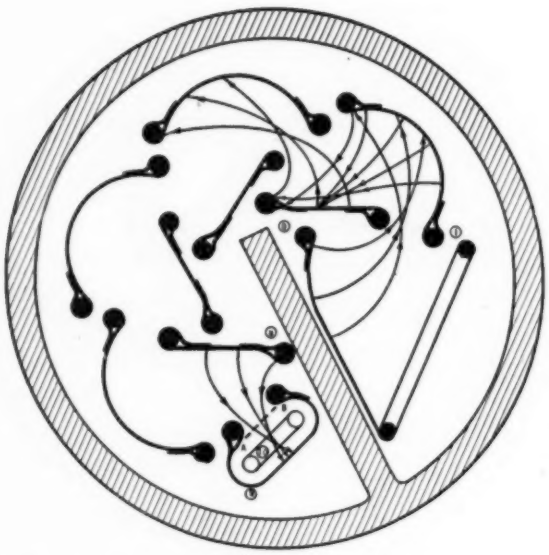


FIG. 1. Internal structure of the RCA 931 photomultiplier tube. The photoelectrons from (0) are drawn over to (1), the first dynode. The secondary electrons leaving (1) are drawn to the next dynode, and so on through nine stages of multiplication, the last occurring at dynode (9). (10) is a double grid serving as the collector. It is almost entirely enclosed by (9). The dashed line AB represents the estimated position of an equipotential joining on to (9). It is assumed in our calculation that only when an electron is in the space enclosed by (9) and AB does an appreciable fraction of its lines of force end on (10).

the mean square fluctuation voltage

$$\langle v^2\rangle_{Av}-\langle v\rangle_{Av}^2=2e\langle i(t)\rangle_{Av}\frac{\langle M^2\rangle_{Av}}{\langle M\rangle_{Av}}$$

$$\times\int_{\nu=0}^{\infty}\langle|T\mathbf{B}(\nu)|^2\rangle_{Av}\cdot|Z(\nu)|^2\cdot d\nu. \quad (5a)$$

When the impedance differs from zero only in the low frequency region where (4a) applies,

$$\langle v^2\rangle_{Av}-\langle v\rangle_{Av}^2=2e\langle i(t)\rangle_{Av}\frac{\langle M^2\rangle_{Av}}{\langle M\rangle_{Av}}\int_{\nu=0}^{\infty}|Z(\nu)|^2\cdot d\nu. \quad (5b)$$

Setting $M=1$ gives the well-known expressions for the saturated diode.

Our problem is seen to be that of the calculation of $\mathbf{B}_k(\nu)$. Now, in the kth burst, the time of arrival, t_i, of each elementary pulse is distributed about a mean value $\langle t_i\rangle_k$ according to some distribution function for the spread in time of arrival, $f_k(t_i)$, which we normalize to unit area. Since M_k is a very large number (over the range of 931's and usual operating conditions it lies between 3×10^4 and 3×10^6) we may write

$$B_k(t)=\int_{t_i=-\infty}^{t}\Delta_k(t-t_i)\cdot f_k(t_i)dt_i, \quad (6)$$

on condition that it is permissible to replace the Δ_i's by Δ_k, a mean shape for the elementary pulses making up the kth burst. If use of this mean shape is legitimate, we may also regard t_i in (6) as the time at which the elementary pulse starts, and $f_k(t_i)$ as the distribution function for time of arrival at the starting point.

The applicability of the assumption in the case of the 931 is discussed in Section III.

Since $B_k(t)$ has the form (6), we can apply a well-known theorem in Fourier analysis[3] to obtain: *The Fourier transform* $\mathbf{B}_k(\nu)$ *of the burst shape is the product of the Fourier transform of the elementary pulse shape and the Fourier transform of the distribution curve for time occurrence of the elementary pulses.* Thus we can consider separately the effects of finite velocity of the electrons and spread in their time of arrival.

III. APPLICATION TO THE RCA 931

The RCA 931[7] is an electrostatic multiplier with nine stages of multiplication. Figure 1 shows the structure of the tube, and some of the electron trajectories.[8] Electrode (0) is the photosensitive cathode. The photoelectrons are drawn to electrode (1), the first multiplying electrode or "dynode," where the process of multiplication starts. It continues through nine stages. The arrangement of the ninth dynode and the anode (electrode (10)) should be noted. The anode is a double grid almost entirely enclosed by (9). The electrons from (8) follow the trajectories shown, passing through the imaginary surface AB (an equipotential joining on to (9)), then through the Faraday cage bounded by the two

[7] RCA has replaced the 931 by the 931-A, which is the same tube made to closer tolerances, and has also put on the market other versions of the tube with the same geometry but different sensitivity and power rating. The 1P21 also uses the "S4" surface, but has higher gain and lower power rating. The 1P22 uses the "S8" surface, the 1P28 the "S5." See *RCA Technical Data Sheets* for details of these tubes. The calculations made in this report should apply within the indicated limits to the whole family.

[8] I am indebted to Dr. J. Rajchman of the RCA Laboratories for permission to use this figure.

sides of (10), and thence to (9). This produces a bi-directional current pulse in the anode circuit, positive while the electrons approach (10), zero while they are inside (10), and negative while they are moving away. As the area under it totals zero, it contributes only to the high frequency noise. At (9) the electrons multiply, and the secondaries are drawn back to (10), giving a positive pulse on the heels of the weaker two-sided one. The pulse shape, calculated as described below, is shown in Fig. 2; in this figure m is the multiplication at (9).

Because of the electrode structure, the model used in the derivation of II—where the elementary pulse Δ does not start till all the multiplication has taken place—does not apply exactly. But a slight broadening of the assumption made in Eq. (6) covers the situation. We can regard $f_k(t_i)$ as the distribution function for time of arrival at AB, t_i being that time and $\Delta_k(t-t_i)$ as the average elementary pulse $ABCDEFGHIJ$ of Fig. 2, with m set equal to the average multiplication at (9). It is almost certain that the neglected variations in Δ—due now to

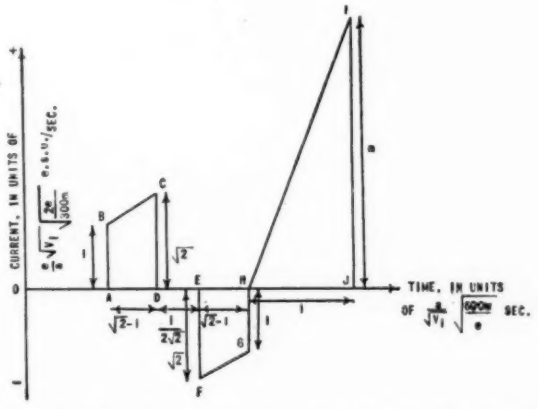


FIG. 2. The elementary current pulse in the external circuit due to the entrance of a single electron into the collecting space. With the inter-stage voltage $V_1=100$ and the distance a (see Fig. 3) 0.085 cm, the unit of time is 2.9×10^{-10} sec. and the unit of current is 1.1×10^{-9} amp. At A, when the electron crosses the imaginary surface AB of Fig. 1, an appreciable fraction of the electron's lines of force begins to terminate on the anode (10). As the electron is accelerated toward the anode, the induced current increases (BC). It falls to zero (DE) when the electron gets into the Faraday cage formed by the two sides of (10). On emerging from the other side of (10), the electron induces a current in the opposite direction, which decreases as the electron is slowed down (FG). On reaching (9) the electron is brought to rest (GH) and m secondaries are produced (with $V_1=100$, m is around 4). These secondaries are accelerated toward (10), giving the forward pulse HIJ, which is terminated when they reach (10) and are collected.

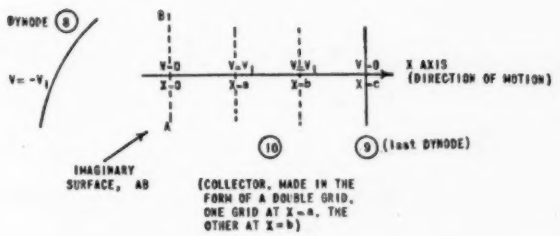

DYNODE (8)
X AXIS (DIRECTION OF MOTION)
IMAGINARY SURFACE, AB
(COLLECTOR, MADE IN THE FORM OF A DOUBLE GRID, ONE GRID AT X=a, THE OTHER AT X=b)
(9) (last DYNODE)

FIG. 3. Equipotentials in the collecting space. The electrons move along the X axis. After crossing AB ($x=0$) they are accelerated toward the anode ($x=a$), coast through it to its other side ($x=b$), and are decelerated as they approach the last dynode (9) at $x=c$. They are stopped there, and the secondary electrons produced are drawn (in the direction of decreasing x) to the anode, where they are collected.

fluctuations in m as well as to time and space spreads between AB and (10)—only begin to affect the spectrum at frequencies so high that the spread in time of arrival at AB and the duration of the average elementary pulse have already reduced the noise output appreciably. The inaccuracy introduced into our calculated spectrum is probably no greater than that caused by variations in performance from one 931 to another, the result of both being to make our results above several hundred megacycles significant only as regards order of magnitude and general trend.

Let us now calculate the average elementary pulse of Fig. 2. It would seem to be a fairly good approximation to assume that only when the electron crosses AB (Fig. 1) does it begin to induce an appreciable amount of charge on the anode. We need, then, to calculate the motion of the electron from this time till the instant it reaches (9), and the motion of its secondaries from the instant they leave (9) till they arrive at (10). It is seen from Fig. 1 that the idealization of plane parallel equipotential surfaces with motion perpendicular to them corresponds closely to the actual state of affairs. It is also a good approximately to take $a=b-a=c-b$ (see Fig. 3). In Fig. 3 the potential is measured from that of (9) taken as a reference level. Under usual conditions, space charge is negligible,[7] and the dynamical problem is the simple one of the motion of a single electron in a uniform field.

The initial velocities of the secondary electrons are of the order of 5 volts.[4] As this is only a small fraction of the velocity acquired in falling to the

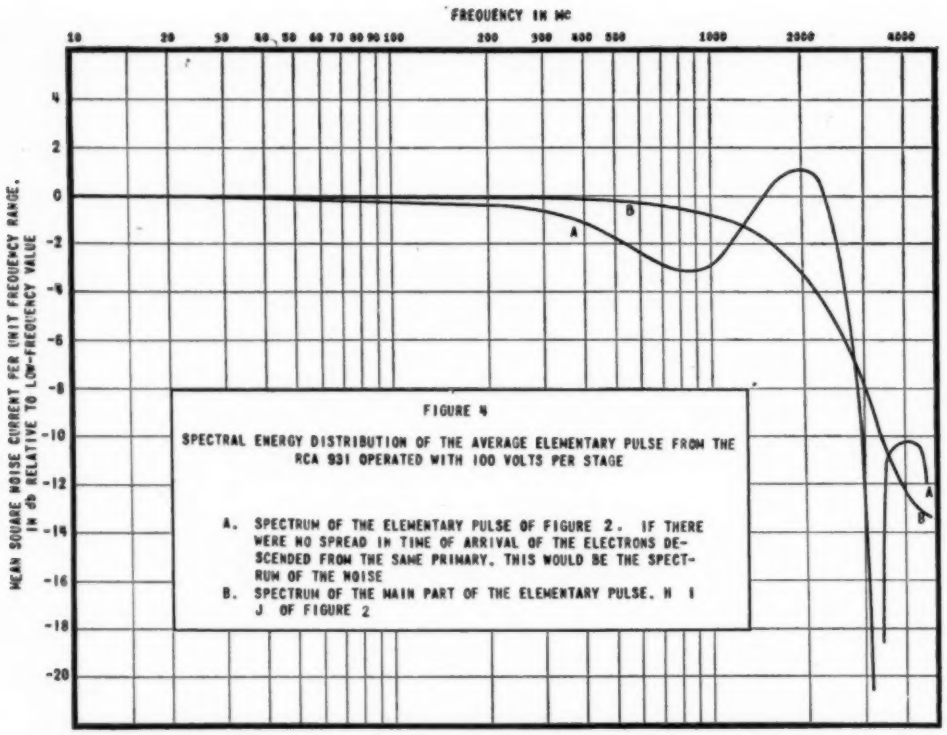


FIG. 4. Power spectrum of the average elementary pulse in the RCA 931 operated with 100 volts per stage. Curve A is the power spectrum of the elementary pulse shown in Fig. 2. If there were no spread in the time of arrival of the electrons descended from the same photon, this would be the spectrum of the noise. Curve B is drawn in for comparison. It is the spectrum of the main part of the pulse, HIJ of Fig. 2, considered by itself. This would be the spectrum of the noise from a saturated plane parallel diode with 100 volts potential difference and electrode spacing of 0.085 cm (assuming zero emission velocity).

next electrode, we may neglect the initial velocities of the secondaries from (8) and (9) in the calculation of the average elementary pulse.

Straightforward solution of the dynamical problem gives the position and the velocity as a function of the time. At AB (Fig. 3) the electron has the velocity corresponding to V_1 (5.9×10^8 cm/sec. for $V_1=100$ volts). In going to $x=a$, its velocity increases linearly with the time, reaching a value $\sqrt{2}$ times that at AB. It then coasts at the same speed from $x=a$ to $x=b$. Between $x=b$ and $x=c$, the velocity decreases linearly to the value corresponding to V_1. The secondaries made at $x=c$ are then accelerated back to $x=b$, acquiring the velocity corresponding to V_1.

With $V_1=100$ volts and $a=0.085$ cm, the total time elapsed is 6.3×10^{-10} second.

As to the electrostatic problem: the distribution of the lines of force of a point charge between two infinite parallel conducting planes has been worked out by Maxwell in his *Treatise*.[9] With the planes a distance d apart, the induced charge on one plane is ey/d, where e is the value of the point charge and y is its distance from the other plane. It follows that the induced current is given by e/d times the velocity.

At $x=a$, $x=b$, and $x=c$ there are conducting planes. The plane $AB(x=0)$ is, however, merely a mathematical surface that is no longer equipotential when an electron is in its vicinity. Our assumption that no lines of force from an electron reach 10 until it has passed AB is, however, equivalent to assuming that AB is a conducting surface, and we shall therefore make no further error by considering it as such. Applying Maxwell's result to our solution of the dynamical problem, we obtain for the current to (10) the

[9] J. C. Maxwell, *Treatise on Electricity and Magnetism* (Oxford University Press, England, 1892), third edition, Vol. 1, p. 266.

result shown in Fig. 2. This is the average elementary pulse. With $V_1=100$ volts and $a=0.085$ cm, the unit of time is 2.9×10^{-10} sec.; the unit of current is 1.1×10^{-9} amp. m is the multiplication at 9. With 100 volts per stage, its average value is around 4; we use $m=4$ in our calculation. Note that while the pulse lasts 6.3×10^{-10} sec., the main part of it is only 2.9×10^{-10} sec. long and has a width between half-amplitude points of only half this.

Curve A of Fig. 4 is the Fourier analysis of this pulse. The ordinate is the mean square noise current per unit frequency range, relative to that at low frequency, in decibels (db). The peculiar shape of the elementary pulse accounts for the rise around 1900 Mc, where the positive and negative portions of the pulse are in resonance. The main part of the pulse, HIJ of Fig. 2, has, by itself, the spectrum shown in curve B.[10]

At 254 Mc ($2\pi\nu\tau=1$, where $\tau=6.3\times10^{-10}$ sec.), the spectrum is down less than 1 db. Though it falls to -3.2 db at 900 Mc, it does not reach -6 db till around 2800 Mc.

If the accelerating voltages between (8) and (9) and (9) and (10) were increased by the same factor, the frequency scale of Fig. 5 would be shrunk by the square root of this factor (i.e., the spectrum would be flat to higher frequencies). Note that only the last two inter-stage voltages are involved; the preceding ones have no effect on the spectrum of the elementary pulse.

We now have to consider the other major cause of frequency dependence, the spread in the time of arrival of the electrons at AB (Fig. 1). This may be because of: (1), fluctuating time delays in secondary emission and (2), differences in the initial velocities of the secondary electrons. As a consequence of (2), the secondaries liberated by a particular electron will take different times to reach the next dynode. They will also arrive there at different points; this will lead to a spatial spreading of the cascade, giving different times of flight for the different branches. Little is known about (1), other than that spread due to it in a five-stage multiplier of design similar to that of the 931 is negligible up to 500 Mc.[4] We shall assume that it only enters at frequencies so high that we can neglect it in our calculation. The spread in emission velocities undoubtedly has a considerable effect. As just remarked, it makes itself felt directly in the time of flight from one point of a dynode to the next dynode and indirectly through the geometrical widening of the cascade. The latter effect is important when one considers the amplification of a high frequency signal,[4] but it should have less bearing on the noise problem. In effect, we are only concerned with the time spread *within* an individual burst, originating with one electron leaving some point of the photo-cathode; its spatial spreading is due to the variations in velocity of the secondaries made in the subsequent stages. As the initial kinetic energies are small relative to the accelerating potential differences, the spread should be small in the first few stages; it should be kept within bounds in the later stages by the concentration of the beam from stage to stage.[4] The only quantitative information that we have on this point has been supplied by Glover, Engstrom, and Pietenpol of the RCA Laboratories, who state that the spread parallel to the axis of the tube is of the order 0.5 cm in the last stage. Dispersion in this direction should have little effect on the time of arrival at the surface AB; and spread in the plane perpendicular to

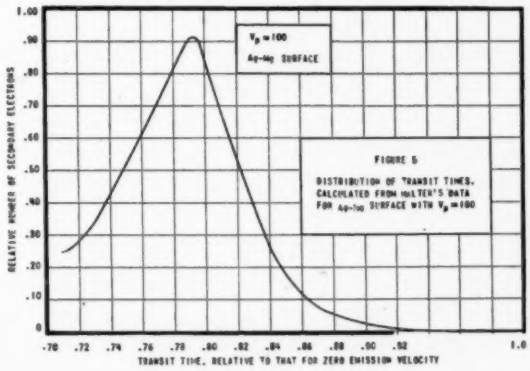


FIG. 5. Distribution of inter-stage transit time. The curve shows the relative number of secondary electrons per unit transit time as a function of the transit time. The transit time is measured in terms of that with zero emission velocity, and is therefore less than or equal to 1. The curve has been computed from Malter's measured distribution of normal emission velocities for the secondaries liberated from an Ag-Mg surface bombarded by 100-volt electrons. It has no quantitative significance for the 931, but should serve in our order of magnitude calculation.

[10] This curve agrees essentially with those calculated for a saturated diode with zero emission velocity by Ballantine, J. Frank. Inst. **206**, 159 (1928) and Spenke, Wiss. Veröff. Siemens [III], **16**, 127 (1937).

the axis should be much less, because of focusing in this plane.

Calculation of the direct effect of spread in emission velocity on inter-stage transit time requires knowledge of the distribution in speed and direction of the secondaries, as well as of the electrostatic field between stages. No data on the velocities from the "S4," "S5," or "S8" surfaces is known to the author; also, it is clear from Fig. 1 that the electron optics of the tube is rather complicated. For the case of plane parallel electrodes, the transit time is determined simply by the normal component of the emission velocity, and Malter has published[4] curves for the distribution of this component, for a silver-magnesium surface bombarded by 100-volt and 200-volt primaries. Figure 5 has been calculated from the curve for 100-volt primaries in his Fig. 3; it shows the number of secondaries per interval of transit time as a function of the transit time, which is measured relative to the maximum possible one (zero initial velocity). What with the directionality of secondary emission (cosine law[11]), the focusing in the tube, and the similarity between the Ag-Mg surface and those used in the various forms of the 931, the curve of Fig. 5, should serve in a calculation of the order of magnitude of the expected effect.

The most probable transit time is seen to be 79 percent of the maximum, given, in seconds, by

$$3.37\times10^{-8}(d/V)^{\frac{1}{2}},$$

where d is the distance between the electrodes in cm and V the potential difference in volts. The half-width of the distribution curve is 10.5 percent of the most probable value; this half-width should not be too different from twice the root mean square deviation.

The over-all transit time to AB involves the sum of eight independently fluctuating inter-stage transit times (from (1) to AB). The spread in the time of arrival at AB can therefore be calculated from the statistical theorem which states that the mean square deviation of the sum of a number of independently fluctuating variables is the sum of the mean square deviations of the independent variables. Calculating the

[11] M. Lortie, *Le phénomène physique d'émission secondaire* (CSF-SFR Conferences de Documentation, Paris, May, 1942).

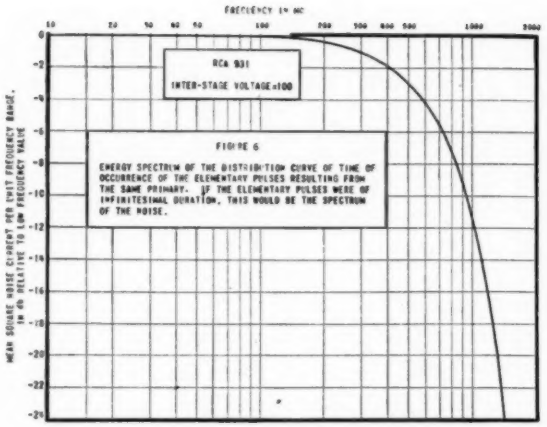

RCA 931
INTER-STAGE VOLTAGE=100
FIGURE 6
ENERGY SPECTRUM OF THE DISTRIBUTION CURVE OF TIME OF OCCURRENCE OF THE ELEMENTARY PULSES RESULTING FROM THE SAME PRIMARY. IF THE ELEMENTARY PULSES WERE OF INFINITESIMAL DURATION, THIS WOULD BE THE SPECTRUM OF THE NOISE.

FIG. 6. "Power" spectrum of the distribution curve of time of occurrence of the elementary pulses resulting from the same photon. This is the Fourier transform of the Gaussian distribution curve for the time of occurrence of the elementary pulses constituting a single burst. If the elementary pulses were of infinitesimal duration, the curve would be the power spectrum of the noise. The curve is drawn for an inter-stage voltage of 100.

stage-to-stage spread as indicated in the preceding paragraph, and applying the theorem, we obtain 6×10^{-10} sec. for the spread in time of arrival at AB of the electrons descended from a common primary.

We note that the spread in transit time from (9) to (10) is only 2.5×10^{-11} second, which is small compared to the duration of the elementary pulse and the spread in time of arrival at AB.

As the spread in time of arrival at AB is the resultant of a number of independent fluctuations (eight) that is large from the point of view of the central limits theorem, its distribution curve is normal:

$$\exp\left(\frac{-(t-t_m)^2}{2(.425\Delta t)^2}\right),$$

where Δt is the half-width and t_m the most probable time of arrival. Its Fourier amplitude is, therefore, proportional to

$$\exp\left(\frac{-(.425)^2}{2}(2\pi\nu\Delta t)^2\right),$$

which is plotted in Fig. 6 with $\Delta t=6\times10^{-10}$ sec. At 265 Mc ($2\pi\nu\Delta t=1$) it is down less than 1 db, but it falls very rapidly, in decibels as the square of the frequency.

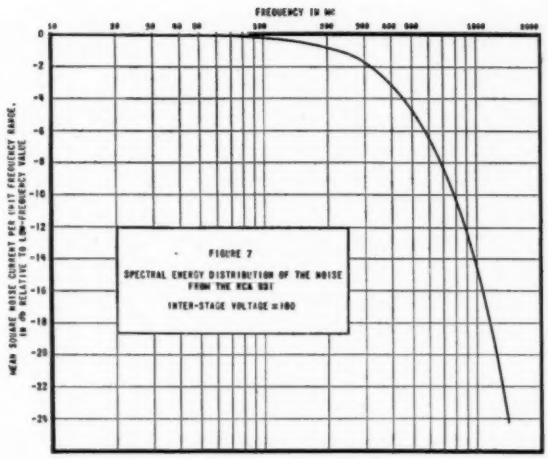


FIG. 7. Power spectrum of the noise from the RCA 931 operated with 100 volts per stage. This curve shows the mean square noise current per unit frequency range *versus* frequency. It is obtained by combining curve *A* of Fig. 4 with the curve of Fig. 6. The rapid fall in the spectrum of the distribution curve of time of arrival of the individual electrons forming the burst, which is caused by the Gaussian character of the distribution, has practically wiped out the rise in the spectrum of the elementary pulse due to a single electron (which is caused by the particular electrode structure).

Figure 5 has as abscissa the transit time relative to that for zero emission velocity. The latter is inversely proportional to the square root of the voltage. Furthermore, Malter's data[4] indicate that the relative spread of emission velocities decreases with increasing primary velocity, so that raising the inter-stage voltages would decrease the spread in transit-time somewhat more than in proportion to the square root of the voltage. Unlike the frequency dependence due to the finite velocities of the electrons, which depends only on conditions in the last two stages, the frequency dependence due to spread in transit time depends on conditions in all the stages from (1) on. Both effects can be pushed up to higher frequencies by increasing the accelerating voltages.

By the theorem in Fourier analysis already referred to, the spectrum of the noise is the product of the two spectra we have just calculated. It is plotted in Fig. 7. At 240 Mc, the output is down 1 db. It reaches 6 db down at 570 Mc, and falls more and more rapidly as the frequency increases. The spread in time of arrival has wiped out, for all practical purposes, the resonance effect due to the two-sidedness of the elementary pulse; the effect of the shape of the spectrum of the elementary pulse is just barely perceptible.

It is clear that no great quantitative significance can be claimed for our final curve, because of the roughness of the data and the method of calculation, as well as the variations in behavior from tube to tube. It does seem safe to state however, that the noise spectrum should be flat up to about 100 Mc, begin to fall off perceptibly in the range 100 to 1000 Mc, and be very weak at higher frequencies. The null at around 3300 Mc would occur in an output which is already very weak.

Some confirmation of these predictions is provided by unpublished measurements of P. Jastram at the Radio Research Laboratory. They showed that the noise output around 30 Mc is about the same as that in the video range. There is also some experimental evidence to indicate that at 100 Mc the output is still about the same, and that above this frequency, it becomes much smaller.

I am grateful to Doctors J. H. Van Vleck and S. A. Goudsmit for illuminating discussions.

The Response of Biased, Saturated Linear and Quadratic Rectifiers to Random Noise*

DAVID MIDDLETON**
Radio Research Laboratory, Harvard University, Cambridge 38, Massachusetts
(Received May 3, 1946)

Spectral distributions and powers associated with the output of biased, saturated linear and quadratic rectifiers are determined according to the method of Rice when the incoming disturbance is random noise. Three classes of input spectra are considered: (I), *broad band noise*, where the central frequency is equal to or less than the spectral width, (II), *semi-broad band noise*, for which the central frequency exceeds the width by a moderate amount, and (III), *narrow band noise*, where the width is much less than the central frequency. Rectification of types I and II yields spectra having roughly the same distribution as that of the incoming waves, but for type III an infinite number of separate noise bands are generated and appear in the output, centered about harmonics of the central frequency. It is found that clipping, whether at the "top" or "bottom" of the incoming wave, always spreads the spectrum and reduces the output power. Further, it is shown that clipping due to cut-off alone produces a greater spectral spread than clipping with saturation in addition, for types I and II, but not necessarily for spectra of type III. Symmetrical clipping for classes I and II yields little broadening of the spectrum, even in extreme cases, and for class III waves, the even-harmonic regions are completely missing except for a d.c. component, and only the odd-harmonic zones appear. The behavior of linear and quadratic rectifiers is qualitatively similar in most cases. *The powers in the d.c. and continuous portions of the output spectrum are shown to be independent of the spectral shape of the incoming noise.* Some examples of extreme clipping are considered, numerous curves are included, and a general analysis is outlined.

I. INTRODUCTION

SINCE the nature of noise problems is basically statistical, one is never able to predict precisely the amplitude of the input or output noise wave at any given instant. Quantitatively the most that can be done in this respect is to determine the probability of occurrence of noise peaks, verify that they follow some law of distribution, and compute the values of the average, the mean-square, and the higher moments. A more far-reaching consequence of the statistical character of the noise lies in the fact that there exists no relationship between the noise power (or mean-square-amplitude) spectrum and the law of the distribution of amplitudes by which one may be obtained from the other. For example, the spectra of clipped† and unclipped noise may be identical, while the instantaneous voltage waves whose power spectra we are comparing are quite different, as may be verified experimentally by observation on a cathode-ray oscilloscope. The former disturbance will show a definite ceiling where clipping has occurred, while the latter will be jagged and without a definite boundary. There is no bridge between the average and the instantaneous in problems of this type because information about the phases of the components is always lacking when we seek to pass from the spectral to the instan-

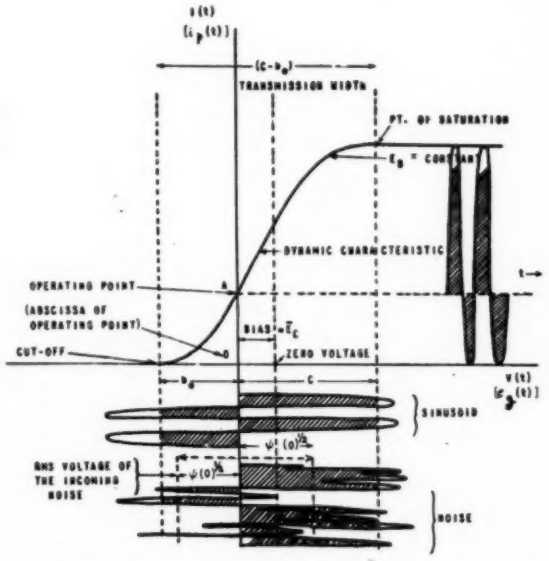


FIG. 1. General dynamic characteristic for resistive loads.

* This paper is based on work done for the Office of Scientific Research and Development under Contract No. OEMsr-411 with the President and Fellows of Harvard College.

** Now located at the Research Laboratory of Physics, Harvard University, Cambridge 38, Massachusetts.

† By clipping we mean amplitude-limiting, either at the top or bottom of the wave or both.

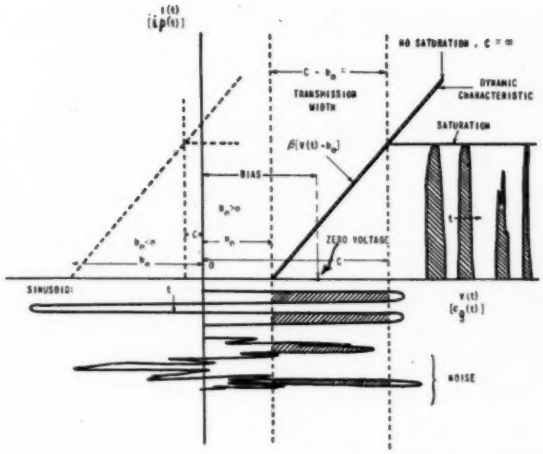


FIG. 2a. Dynamic characteristic for the biased and saturated linear rectifier.

taneous amplitude distribution. Thus for a given probability distribution of amplitudes any number of differently shaped spectra are possible, and *vice versa*.

Now when random noise enters a non-linear device the output disturbance is distorted, and a change in the amplitude distribution and spectrum of the incoming wave results. Cross-modulation, or mixing, of the various components occurs, so that the outgoing wave has generally quite different instantaneous and average characters from those of the incoming signal. For example, the statistics of the amplitude distribution are modified, the noise is no longer "pure" or random in the above sense, and the power and the power spectrum of the output in many instances differ noticeably from those of the input. The factors determining the character of the outgoing wave are chiefly (1), the rectifier response, which is the dynamic $i_p - e_g$ characteristic of the tube in question, (2), the spectrum of the incoming random noise, and (3), the cut-off and saturation levels, which specify the extent of the clipping, as illustrated for the general case in Fig. 1, and specifically for the linear and a general form of quadratic rectifier in Figs. 2a and 2b.

The form of the dynamic response, and to a less extent the output current at saturation, depend also on the load in the plate circuit of the rectifier. The analysis as presented here applies for purely resistive loads, and is a good approximation for more general impedances, provided the reactance X is small compared with the resistance R. For appreciable reactances, however, the analysis becomes much more intricate.

Our rectifier is operated at some bias such that the cut-off is b_0 volts and saturation occurs at C volts, both measured from the abscissa of the operating point 0, as indicated in Fig. 1, where C always exceeds b_0. The (voltage) *transmission width* of the device is defined as that region of the response curve which will pass a portion of the incoming wave. Analytically, as can be observed from Figs. 1 and 2, the transmission width is $C-b_0$ for the linear detector, and $2(C-b_0)$ for the quadratic response. For our "linear" device the output current, or voltage measured across an impedance, is then

$$\begin{aligned} I(t) &= \beta(C-b_0), && C \leqslant V, \\ &= \beta(V-b_0), && V \leqslant C, \\ &= 0, && V \leqslant b_0, \end{aligned} \qquad (1\text{-}1)$$

vide Fig. 2a, where the quantity β is the (dynamic) tube transconductance. When the curvature of the I–V characteristic is small in the vicinity of cut-off and saturation, the linear dynamic path is a fair approximation for a large class of actual operating conditions.

We distinguish two types of quadratic response, which we designate as Type A and Type B. The Type A characteristic, similar to the Type B except that it levels off abruply at $V=C$, Fig. 2b, is less important and less general than the Type B, and so will not be considered further in this paper. The Type B characteristic, obtained analytically by fitting two parabolas together in the manner indicated in Fig. 2b, is

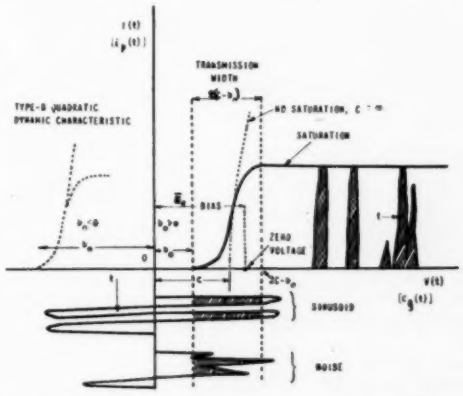


FIG. 2b. Dynamic characteristic for the biased and saturated Type-B quadratic rectifier.

perhaps the simplest theoretical representation of a general form of dynamic response often encountered in practice, particularly in that curvature in the neighborhood of cut-off and saturation is considered. Then for Type B tubes we may write the output current as

$$\left.\begin{aligned} I(t) &= 2\lambda^2(C-b_0)^2, \quad 2C-b_0 \leqslant V, \\ &= \lambda^2\{2(C-b_0)^2-(V-2C+b_0)^2\}, \\ &\qquad\qquad C \leqslant V \leqslant 2C-b_0, \\ &= \lambda^2(V-b_0)^2, \quad b_0 \leqslant V \leqslant C, \\ &= 0, \quad V \leqslant b_0, \end{aligned}\right\} \quad (1\text{-}2)$$

where now λ is a scale factor with the dimensions of mhos.

The purpose of this paper is to investigate both qualitatively and quantitatively the effects of cut-off and saturation in determining the output spectra and powers when random noise enters a non-linear device similar to that considered above, shown in Fig. 2. The theory is developed following a method originally suggested by S. O. Rice of the Bell Telephone Laboratories.[1] The assumption is made that *initially* the noise is random, with a normal or Gaussian distribution of amplitudes. In practice this is never precisely true, as the larger noise peaks are inevitably limited to some extent by nominally linear circuit elements such as amplifiers. However, when the limiting occurs but a small percentage of the time, our "pure" noise may then be substituted for the actual disturbance with little error and the results so obtained apply in the practical case.

II. REMARKS ON THE METHOD OF ANALYSIS

To calculate the power and spectra for the noise[1–3] after it has passed through a non-linear device we use the contour representation [1,3–6] of the output current $I(t)$ as a function of the input voltage $V(t)$, namely

$$I(t)=\frac{1}{2\pi}\int_{\mathbf{C}} f(iz)e^{iV(t)z}dz, \quad (2\text{-}1)$$

where $\mathbf{C}$ is a contour extending from $-\infty$ to $+\infty$ along the real axis and is indented downward about a possible singularity at the origin. The function $f(iz)$ is the Fourier transform of the rectifier characteristic and usually has a simple pole or branch-point at $z=0$.

Now if the correlation function $R(t)$ of the rectifier output is known we may calculate the power spectrum $W(f)$ from the important relation, deduced originally by Wiener,[7] and later by Khintchine,[8] and others,[9] *viz*:

$$W(f)=4\int_0^\infty R(t)\cos\omega t dt, \quad \omega=2\pi f, \quad (2\text{-}2)$$

which has the transform

$$R(t)=\int_0^\infty W(f)\cos\omega t df. \quad (2\text{-}3)$$

It is also noted that the mean total power is simply $R(0)$. For an input disturbance consisting solely of noise the correlation function for the output takes the form[10]

$$R(t)=\frac{1}{4\pi^2}\int_{\mathbf{C}} f(iz)\exp(-\psi z^2/2)dz \times\int_{\mathbf{C}'} f(i\xi)\exp(-\psi\xi^2/2-\psi(t)z\xi)d\xi, \quad (2\text{-}4)$$

where $\psi=\psi(0)$ is the mean square input voltage and $\psi(t)$ is the input correlation, having associated with it, in virtue of (2–2), the input spectrum $w(f)$.

III. OUTPUT POWER AND SPECTRA

If, after substituting the value of $R(t)$ from (2-4) in (2-2), we expand $e^{-z\xi\psi(t)}$ in powers of z and ξ, we may write the spectrum $W(f)$ of the

[1] S. O. Rice, "Mathematical analysis of random noise," Bell Sys. Tech. J. **23**, 282 (1944); **24**, 46 (1945).
[2] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, No. 2 and 3, 323 (1943).
[3] D. O. North, "The Modification of Noise by Certain Non Linear Devices," a synopsis of which was presented at the Jan. 1944 Winter Technical Meeting of the I.R.E.
[4] W. R. Bennett and S. O. Rice, "Note on methods of computing modulation products," Phil. Mag. **18**, 422 (1934).
[5] W. R. Bennett, "Response of a linear rectifier to signal and noise," J. Acous. Soc. Am. **15**, 164 (1944). (See also Bell Sys. Tech. J. **23**, 97 (1944).)
[6] K. Fränz, Zeits. f. Hoch. u. Eleck. **57**, 146 (1941), and Eleck. Nach. **17**, 215 (1940); **19**, 285 (1942).
[7] N. Wiener, Acta. Math. **55**, 117 (1930).
[8] Khintchine, Math. Ann. **109**, 604 (1934).
[9] See footnote 12 of reference 3, and also reference 1 Part II.
[10] This is a special case of Eq. (4.8–7) reference 1, in which $g_n(u, v, \tau)$ is set equal to unity.

TABLE I. Comparison of the precise and series* values of the total power output, for the linear rectifier.

	$C/\psi^{\frac{1}{2}}=0.5$			$=1.0$			$=1.5$			$=2.0$			$=\infty$		
$b_0/\psi^{\frac{1}{2}}$	*7 Terms	*14 Terms	$R(0)$	*7 Terms	*14 Terms	$R(0)$	*7 Terms	*14 Terms	$R(0)$	*7 Terms	*14 Terms	$R(0)$	*7 Terms	*14 Terms	$R(0)$
−2.0	3.794	3.795	3.795	4.419−	4.419	4.419	4.765	4.766	4.766	4.919	4.920−	4.921	4.994	4.994	4.994
−1.5	2.222	2.225	2.226	2.731	2.732	2.735	3.026	3.028−	3.028	3.161−	3.162−	3.162	3.228−	3.228	3.228
−1.0	1.108	1.112	1.122	1.505	1.507	1.516	1.751	1.753	1.755	1.867	1.869	1.869	1.842	1.924	1.925
−0.5	.417	.422	.435	.701	.706	.715	.895	.898	.900	.991+	.996	.996	.942	1.055	1.05(6)
0.0	.0867	.0899	.0926	.251	.256	.258	.385	.388	.389	.458	.459	.460	.497	.499	.500
0.4	.0029	.0031−	.0032	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.0	0.0	0.0	.0468	.0490	.0510	.122	.127−	.128	.174	.178−	.178	.207−	.209	.210
0.9	—	—	—	.0014−	.0015	.0017	—	—	—	—	—	—	—	—	—
1.0	—	—	—	0.0	0.0	0.0	.0201	.0217	.0232	.0488	.0517	.0526	.0733	.0746	.0753
1.4	—	—	—	—	—	—	.00055	.00065	.0007	—	—	—	—	—	—
1.5	—	—	—	—	—	—	0.0	0.0	0.0	.0073	.0080	.0086	.0220	.0224	.0228
1.9	—	—	—	—	—	—	—	—	—	.00018	.0002−	.0002−	—	—	—
2.0	—	—	—	—	—	—	—	—	—	0.0	0.0	0.0	.0053	.0056	.0058

$R(0)$ is in units of $\beta^2\psi(0)$ and is obtained from the exact, closed form.

* The series value is obtained by evaluation of $R(0) \doteq \sum_{n=0}^{N} \frac{\psi(0)^n}{n!} h_{0n}^2(-1)^n$ where N is theoretically infinite, but in practice for these computations is 14. Calculations for $N=7$ are also given above.

output in a form convenient for calculation:

$$W(f)=\sum_{n=0}^{\infty}(-1)^n c_{0,n}(f)h_{0,n}^2/n!, \qquad (3\text{-}1)$$

where

$$c_{0,n}(f)=4\int_0^{\infty}\psi(t)^n\cos\omega t dt, \qquad (3\text{-}2)$$

and

$$h_{0,n}=\frac{1}{2\pi}\int_C z^n f(iz)\exp\left[-\psi(0)z^2/2\right]dz. \qquad (3\text{-}3)$$

Note that when $n=0$ we obtain the d.c. power. Now from (2-2) and (3-1)–(3-3) the correlation of the output is

$$R(t)=\sum_{n=0}^{\infty}(-1)^n h_{0,n}^2\psi(t)^n/n!, \qquad (3\text{-}4)$$

the d.c. power becomes

$$W_{\text{d.c.}}=h_{0,0}^2, \qquad (3\text{-}5)$$

while the power in the continuum of the transmitted wave may be expressed by

$$W_c=\sum_{n=1}^{\infty}(-1)^n\psi(0)^n h_{0,n}^2/n!. \qquad (3\text{-}6)$$

We see from (3-6) that *the power in the continuum, and for that matter in the d.c. component of the output disturbance, is independent of the spectral distribution of the input wave,* and for a given current-voltage characteristic depends only on the mean-square input-noise voltage. This result is a somewhat special case of a more general theorem; (see Eq. (4.9-19) of reference 1).

Unfortunately, the series (3-6) does not converge rapidly except when the output spectrum consists of separate spectral regions, a condition which arises if the input wave is narrow-band. In some cases we may use other methods of obtaining the mean power in the continuum. For the total mean output power we have

$$R(0)=W_r=(2\pi\psi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}I(V)^2\exp\left(-V^2/2\psi\right)dV, \qquad (3\text{-}7)$$

with $I(V)$, the output current in a unit resistance, given by (1-1) or (1–2). If we normalize V by the substitution $V=x\psi^{\frac{1}{2}}$, Eq. (3-7) may be written in the more useful form

$$R(0)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}I(\psi^{\frac{1}{2}}x)^2\exp\left(-x^2/2\right)dx. \qquad (3\text{-}8)$$

Equation (3-8) yields the total output power, and the power associated with the continuum may be found from (3-5) by subtraction, since

$$W_c=W_r-W_{\text{d.c.}}=R(0)-h_{0,0}^2. \qquad (3\text{-}9)$$

Note that we may also obtain the d.c. power from

$$W_{\text{d.c.}}=\left\{(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}I(x\psi^{\frac{1}{2}})\exp\left(-x^2/2\right)dx\right\}^2. \qquad (3\text{-}10)$$

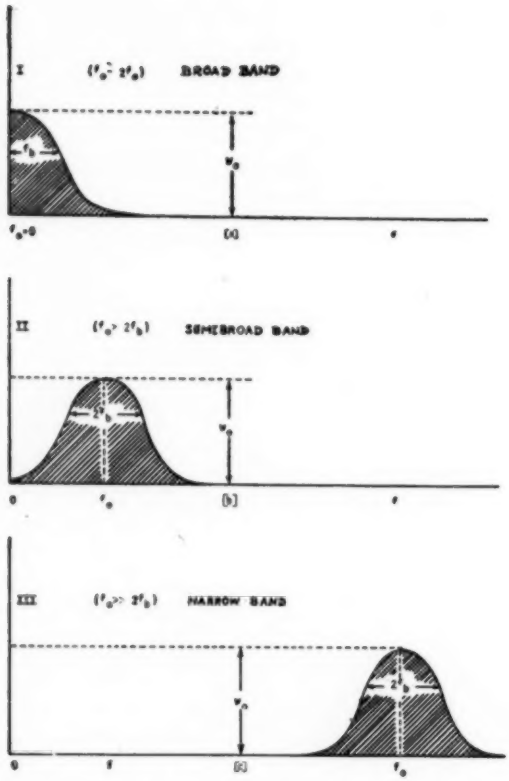


FIG. 3. The three classes of input spectra.

Table I illustrates the numerical difference between the series computation which uses (3-6) and the "precise" value (3-8).

It is convenient to classify input spectra roughly in three groups: (I) *broad band noise*, which is characterized by a band width as large as or greater than the central frequency f_0 of its spectrum, (II) *semi-broad band noise*, where the central frequency is of the same order of magnitude as the spectral width, but always larger, and finally (III), *narrow-band noise*, for which the central frequency is decidedly greater than the band width. Figure 3 illustrates I–III. It is clear that groups I–III may overlap; the extent to which overlapping occurs depends on our choice of what constitutes "band width." To facilitate the calculations in some cases, and to preserve a reasonably clear-cut distinction between classes I-III, we set as our spectral limits determining the band width the frequencies at which the (power) spectral level falls to a few percent of its maximum value. Our classification depends of course, on the assumption that the input spectra have but one maximum, a condition quite common in practice. The analysis is perfectly capable of handling more involved spectral shapes, but the computational difficulties are then greatly increased.

The input spectra considered specifically in this paper have the general form

$$w(f) = W_0 \exp\left[-(\omega-\omega_0)^2/\omega_b^2\right], \quad \omega \geq 0, \quad (3\text{-}11)$$

where $\omega_0 = 2\pi f_0$ is the angular central frequency, ω_b is an angular frequency relating the width to the central frequency, and hence governing the relative spread of the input spectrum, and W_0 is the maximum spectral density, in units of volts^2/frequency. Our choice of (3-11) is based on the observation that the Gaussian shape is a simple and reasonably good approximation of many spectra encountered in practice, particularly for disturbances passing through single or cascaded tuned circuits, in receivers, transmitters, and similar apparatus. Actually, as Wallman[11] has shown, there are no physical filters (i.e., frequency-discriminating devices) that can be constructed with a finite number of passive elements to give precisely a Gaussian response, for only with an infinite time delay is it possible to achieve the rapid falling-off of the response that distinguishes (3-11) for frequencies moderately well away from f_0, for, let us say, $f_0/f_b > 2$. Spectra of the type (3-11), however, appear to be closer approximations in practice than the so-called "ideal" or uniform band.

In what follows we shall find it more convenient to use the normalized input correlation function $r(t)$, rather than $\psi(t)$. Here we have

$$r(t) \equiv \psi(t)/\psi(0) = \int_0^\infty w(f)\cos \omega t df \Big/ \int_0^\infty w(f) df; \quad (3\text{-}12)$$

then $r(0)$ is unity and $r(t) < 1$, $t > 0$. The input correlation for class I spectra ($f_0 \doteq 0$), is easily deduced from (3-11) and (3-12). We have

$$r(t)_{\mathrm{I}} = \exp\left(-\omega_b^2 t^2/4\right); \quad \psi(0) = W_0\omega_b/4\pi^{\frac{1}{2}}. \quad (3\text{-}13)$$

When $f_0 > 0$, $r(t)$ becomes

[11] H. Wallman, "Realizability of filters," Radiation Laboratory Report 637, M. I. T., Dec. 8, 1944.

$$r(t)_{II,III}=(\omega_b\pi^{\frac{1}{2}})^{-1}\int_0^\infty \exp\left[-(\omega-\omega_0)^2/\omega_b^2\right]\cos\omega t\,df$$

$$=(2\pi^{\frac{1}{2}}\omega_b)^{-1}\int_{-\omega_0}^\infty \exp\left(-\omega'^2/\omega_b^2\right)\cos(\omega'+\omega_0)t\,d\omega', \quad (3\text{-}14)$$

where now

$$\psi(0)=W_0\omega_b/2\pi^{\frac{1}{2}}.$$

If there is an ignorable "tail" at $f=0$, we may extend the lower limit of integration to $-\infty$ with negligible error and obtain

$$r(t)_{II,III}=\cos\omega_0 t\,\exp\left(-\omega_b^2t^2/4\right)$$

$$\equiv r_0(t)\cos\omega_0 t, \quad (3\text{-}15)$$

which defines $r_0(t)$, a useful quantity in later work. Note from (3-14) that $\psi_{II,III}=2\psi(0)_I$ since both sides, $f\gtrless f_0$, contribute to the mean power. We shall consider here only examples of class II spectra for which (3-15) applies.

When the incoming wave is a narrow band and has the spectrum (3-11), Eq. (3-14) may be used with even less error than occurs for the semi-broad band case, since here f_0 is very much larger than f_b.

To determine the output spectrum for (I), we see from (3-2) and (3-13) that

$$c_{0,n}(f)_I=4\psi^n(\pi/n)^{\frac{1}{2}}\frac{\exp\left(-\omega^2/n\omega_b^2\right)}{\omega_b}. \quad (3\text{-}16)$$

Similarly from (3-2) and (3-15) the semi-broad band case yields

$$c_{0,n}(f)_{II}=4\psi^n\int_0^\infty \exp\left(-n\omega_b^2t^2/4\right)$$

$$\times(\cos\omega_0 t)^n\cos\omega t\,dt, \quad (3\text{-}17)$$

and this may be evaluated with the help of

$$(\cos\omega_0 t)^n=2^{1-n}\sum_{j=0}^{n/2,\,(n-1)/2}{}_nC_j\,{}_n\Delta_j\cos(n-2j)\omega_0 t, \quad (3\text{-}18)$$

where

$${}_n\Delta_{n/2}=\tfrac{1}{2},\quad {}_n\Delta_j=1;\quad n\neq 2j,\quad \text{and}\quad {}_nC_j=n!/(n-j)!j!$$

is the familiar binomial coefficient. The limits on the summation apply for even or odd values of n, respectively. Substitution of (3-18) in (3-17) leads to integrals of the type

$$\int_0^\infty \exp\left(-n\omega_b^2t^2/4\right)\{\cos[(n-2j)\omega_0+\omega]t$$

$$+\cos[(n-2j)\omega_0-\omega]t\}\,dt,$$

which appears to be resolved into components representing separate spectral zones distributed about $(n-2j)f_0$. However, since f_b and f_0 are comparable in magnitude, one "harmonic" zone $l=l'$, corresponding to all values of $(n-2j)$ which are equal to l, overlaps another zone, $l=l''$, etc., so that physically the output spectrum cannot be resolved into distinct harmonic noise bands. It is represented by the superposition of all the different zones $l\geq 0$. We may apply (3-18) nevertheless, to obtain finally

$$c_{0,n}(f)_{II}=4\psi^n(\pi/n)^{\frac{1}{2}}\frac{1}{\omega_b}K_n(\omega/\omega_b)$$

$$\times\exp\left(-\omega^2/n\omega_b^2\right), \quad (3\text{-}19)$$

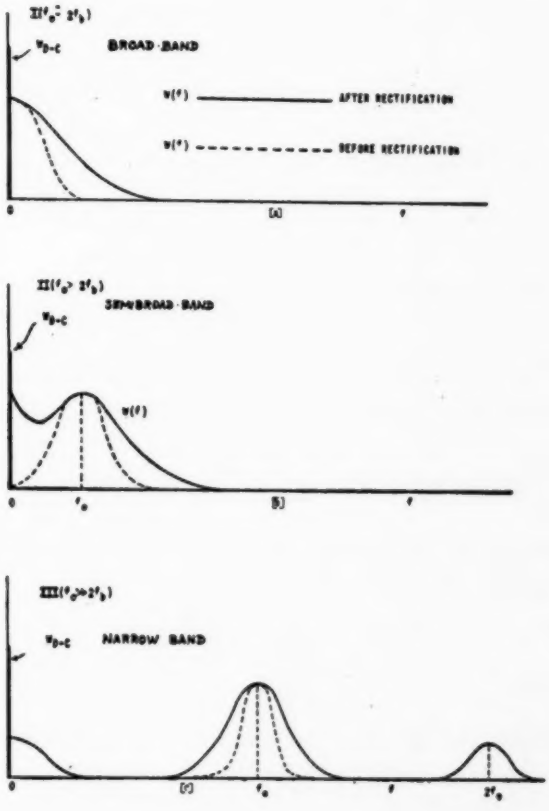


FIG. 4. The three classes of input spectra after rectification, arbitrary normalization.

where

$$K_n(\omega/\omega_b)=2^{1-n}\sum_{j=0}^{n/2,\,(n-1)/2}{}_n\Delta_j\,{}_nC_j \times\cosh\left[\frac{2(n-2j)\omega_0\omega}{n\omega_b^2}\right]\times\exp\left[-(n-2j)^2\omega_0^2/n\omega_b^2\right]. \quad (3\text{-}20)$$

Spectral "tails" at $f=0$ are ignored, as they are relatively insignificant.

When narrow band noise (III) enters our non-linear device, a separation of the harmonic zones becomes physically possible, since now $f_0\gg f_b$. By the substitution of Eqs. (3-18) in (3-17) we are able to pick out the various spectral regions analytically; in practice we use filters having suitable response and centered about lf_0.* For a given value of l we require all n for which $j=(n-l)/2\geq 0$, where n is odd for odd l, and even for even l. Then (3-17) may be modified to give

$$c_{0,\,n}{}^l(f)_{\mathrm{III}}=\left[4\psi^n n!\,{}_n\Delta_{(n-l)/2}\Big/\left(\frac{n-l}{2}\right)!\left(\frac{n+l}{2}\right)!2^{n-1}\right]\int_0^\infty r_0(t)^n\cos\omega t\cos l\omega_0 t\,dt$$

$$=\left[4\psi^n n!\,{}_n\Delta_{(n-l)/2}\Big/\left(\frac{n-l}{2}\right)!\left(\frac{n+l}{2}\right)!2^{n-1}\right]\times\{\exp\left[-(l\omega_0+\omega)^2/n\omega_b^2\right]+\exp\left[-(l\omega_0-\omega)^2/n\omega_b^2\right]\}/2. \quad (3\text{-}21)$$

Because $lf_0\gg f_b$, we may safely ignore the term $\exp\left[-(l\omega_0+\omega)^2/n\omega_b^2\right]$, except when $l=0$. Figure 4 illustrates the form of the spectral output for classes I–III. The normalization of the spectra is arbitrary.

By taking $l=n=0$ we obtain the d.c. output power,

$$W_{\mathrm{d.c.}}=h_{0,\,0}{}^2,\quad\text{since}\quad\int_0^\infty c_{0,\,0}{}^0(f)\,df=2\int_0^\infty\delta(f-0)\,df=1. \quad (3\text{-}22)$$

The correlation function corresponding to the lth zone is

$$R_l(t)=\epsilon_l\cos l\omega_0 t\sum_{n=l,\,l+2,\cdots}^{\infty}(-1)^n\psi^n h_{0,\,n}{}^2 r_0{}^n(t)\Big/\left(\frac{n-l}{2}\right)!\left(\frac{n+l}{2}\right)!2^n, \quad (3\text{-}23)$$

where

$$\epsilon_l=2,\quad l\geq 1;\quad \epsilon_0=1,\quad\text{and}\quad r_0(t)=\exp\left(-\omega_b{}^2t^2/4\right).$$

When we take $t=0$ it follows that the mean power in the low frequency continuum is determined provided $n=2,\,4,\,\cdots,\,l=0$:

$$W_{\mathrm{l.f.}}=\sum_{2,\,4\cdots}^{\infty}\psi^n h_{0,\,n}{}^2/2^n(n/2)!^2. \quad (3\text{-}24)$$

Equation (3-23) also enables us to find the power associated with each of the higher ($l\geq 1$) spectral regions. Setting $t=0$ in (3-23) gives

$$W_l=\sum_{n=l,\,l+2,\cdots}^{\infty}(-1)^n\psi^n h_{0,\,n}{}^2\Big/2^{n-1}\left(\frac{n-l}{2}\right)!\left(\frac{n+l}{2}\right)!,\quad l\geq 1. \quad (3\text{-}25)$$

Note that W_l is independent of the spectral distribution of the input wave.[12]

* Note: here we tacitly assume "ideal" filters, with a uniform response in the regions of interest, so that all components of a particular harmonic band are passed without distortion. When this is not the case the output is $W_l(f)|G(f)|^2$, where $G(f)$ is the *amplitude* response of the filter in question.

[12] This is an example of Eq. (4.9-19), reference 1.

IV. THE BIASED, SATURATED LINEAR RECTIFIER

The explicit calculation of powers and spectra depends on a knowledge of $h_{0,n}$, Eq. (3-3). In particular we are interested in the general biased and saturated linear rectifier shown in Fig. 2, for which the Fourier transform, i.e., $f(iz)$, of the i_p-e_g characteristic (Eq. 1-1) is readily demonstrated to be

$$f(iz)=\beta[\exp(-ib_0z)-\exp(-iCz)]/(iz)^2, \quad C\geq b_0, \tag{4-1}$$

so that (3-3) becomes

$$h_{0,n}=-\frac{\beta}{2\pi}\int_C z^{n-2}[\exp(-ib_0z)-\exp(-iCz)]\exp(-\psi z^2/2)dz, \quad C\geq b_0. \tag{4-2}$$

This integral is evaluated in the Appendix, and yields the following results:

$$h_{0,0}=-\beta\psi^{\frac{1}{2}}/2\{(\bar{C}\Theta_c-1)-\bar{b}(\Theta_{b_0}-1)+2(\phi_c-\phi_{b_0})\}, \tag{4-3}$$

$$h_{0,1}=-\beta i/2\{\Theta_c-\Theta_{b_0}\}, \tag{4-4}$$

$$h_{0,n}=\beta(-1)^{(n+2)/2}\psi^{(1-n)/2}\{\phi_c^{(n-2)}-\phi_{b_0}^{(n-2)}\}, \quad n=2, 4, 6\cdots, \tag{4-5}$$

$$h_{0,n}=\beta i(-1)^{(n+3)/2}\psi^{(1-n)/2}\{\phi_c^{(n-2)}-\phi_{b_0}^{(n-2)}\}, \quad n=3, 5, 7\cdots. \tag{4-6}$$

In the above, and for subsequent use, we have adopted the abbreviations

$$\bar{b}_0=b_0/\psi^{\frac{1}{2}}; \quad \bar{C}=C/\psi^{\frac{1}{2}}; \quad \Theta_{b_0}=\Theta(b_0/(2\psi)^{\frac{1}{2}}); \quad \Theta_c=\Theta(C/(2\psi)^{\frac{1}{2}}), \tag{4-7}$$

and

$$\phi_c^{(j)}=\phi^{(j)}(C/\psi^{\frac{1}{2}}); \quad \phi_{b_0}^{(j)}=\phi^{(j)}(b_0/\psi^{\frac{1}{2}}), \quad j=0, 1, 2, \cdots.$$

where ϕ is the same as $\phi^{(0)}$. The quantity Θ denotes the familiar error function, here written as

$$\Theta(x_0)\equiv(2/\pi^{\frac{1}{2}})\int_0^{x_0}\exp(-y^2)dy; \quad \Theta(x_0)=-\Theta(-x_0), \tag{4-8}$$

and $\phi^{(j)}$ takes the form

$$\phi^{(j)}(x_0)=\frac{d^{(j)}}{dx^{(j)}}\frac{\exp(-x^2/2)}{(2\pi)^{\frac{1}{2}}}\bigg|_{x=x_0}=(-1)^jH_j(x_0)\exp(-x_0^2/2)/(2\pi)^{\frac{1}{2}}, \tag{4-9}$$

$H_j(x_0)$ being an Hermitian polynomial of the jth order. Extensive curves illustrating the effects of cut-off and saturation on the output spectra after rectification by a biased and saturated linear detector are given in Figs. 5-17,* based on (3-16), (3-19), and (3-21) in (3-1), when $h_{0,n}$ is obtained from (4-3)–(4-6).

In many cases the power output is of interest. Because of the relatively slow convergence of the series (3-4) and (3-6) we use instead the result (3-8) to compute the mean total output power W_τ. For our device, characterized by Eq. (1-1), (or (4-1)), we see that

$$W_\tau=R(0)=\frac{\beta^2\psi}{(2\pi)^{\frac{1}{2}}}\left\{\int_{b_0/\psi^{\frac{1}{2}}}^{\bar{c}}(x-\bar{b}_0)^2\exp(-x^2/2)dx+\int_{\bar{c}}^{\infty}(\bar{C}-\bar{b}_0)^2\exp(-x^2/2)dx\right\}, \tag{4-10}$$

irrespective of whether or not the input spectrum belongs to groups I-III. Equation (4-10) integrates to

$$W_\tau=\beta^2\psi\{(\bar{C}-\bar{b}_0)^2/2+(1+2\bar{C}\bar{b}_0-\bar{C}^2)\Theta_c/2-(1+\bar{b}_0^2)\Theta_{b_0}/2+2\bar{b}_0\phi_c+\phi_c^{(1)}+\phi_{b_0}^{(1)}\}. \tag{4-11}$$

The d.c. output power follows from (3-5) and (4-3), and that in the continuum from (3-9) and the above relation (4-11). Figures 18–20 illustrate the dependence of W_τ, $W_{\text{d.c.}}$, and W_c on $\bar{b}_0$ and $\bar{C}$, while Figs. 21 and 22 show W_l, calculated from (3-24) and (3-25), for $l=0$ and 1.

---

* Note: in the figures we have represented $W(f)$ formally by its equivalent $W(\omega/\omega_b)$.

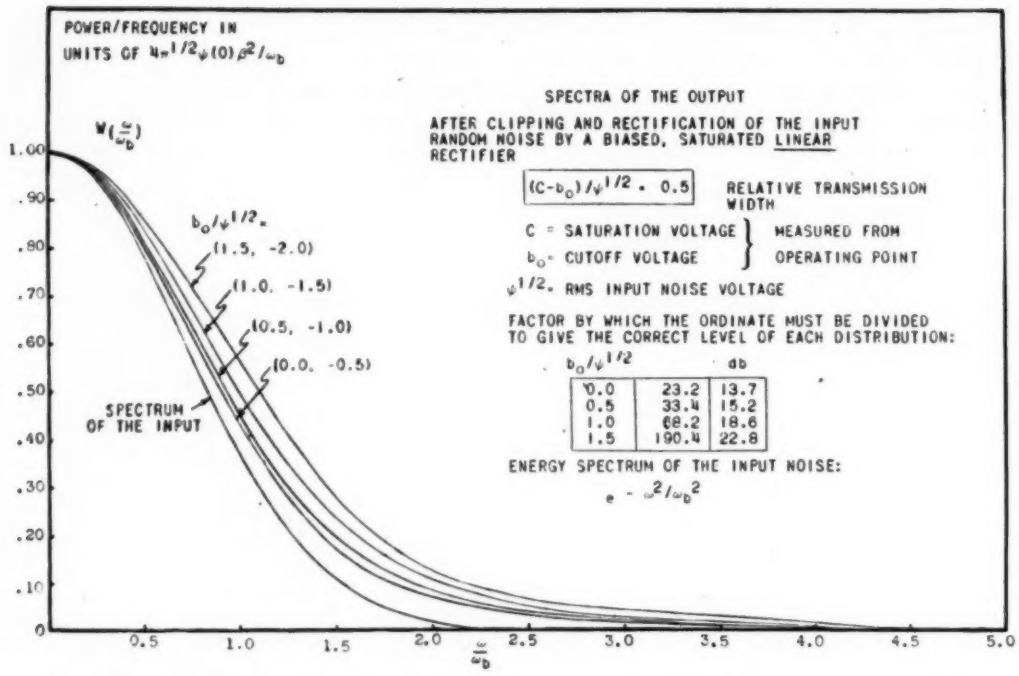


FIG. 5. Output spectra for broad band input noise (I) from a biased and saturated linear rectifier. (In this and succeeding figures, the correct output spectral level is expressed in units of $4\pi^4\psi(0)\beta^2/\omega_b$, *after* division by the factors indicated.)

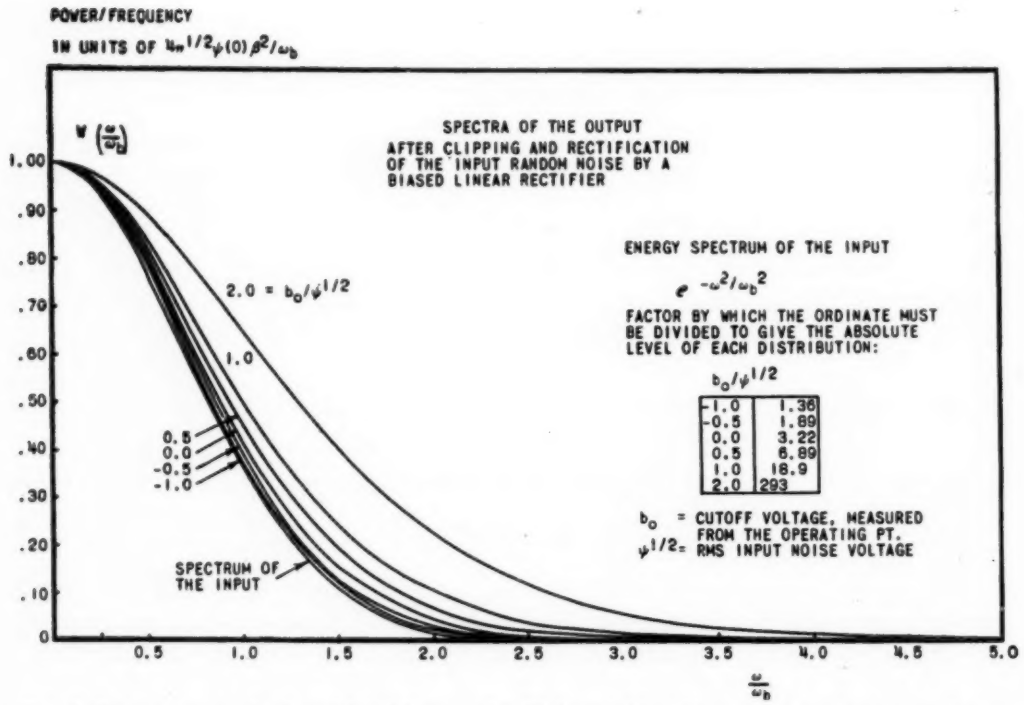


FIG. 6. Output spectra for broad band input noise (I) from a biased but unsaturated linear rectifier.

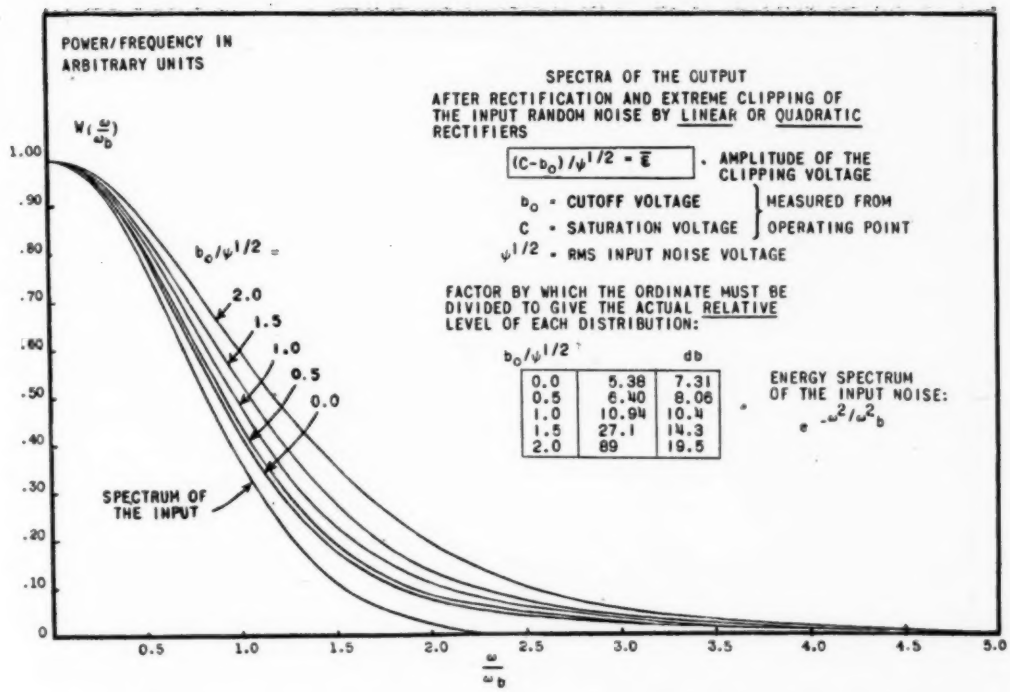

POWER/FREQUENCY IN ARBITRARY UNITS
SPECTRA OF THE OUTPUT
AFTER RECTIFICATION AND EXTREME CLIPPING OF THE INPUT RANDOM NOISE BY LINEAR OR QUADRATIC RECTIFIERS
$(C-b_0)/\psi^{1/2} = \bar{\epsilon}$. AMPLITUDE OF THE CLIPPING VOLTAGE
b_0 = CUTOFF VOLTAGE } MEASURED FROM
C = SATURATION VOLTAGE } OPERATING POINT
$\psi^{1/2}$ = RMS INPUT NOISE VOLTAGE
FACTOR BY WHICH THE ORDINATE MUST BE DIVIDED TO GIVE THE ACTUAL RELATIVE LEVEL OF EACH DISTRIBUTION:
$b_0/\psi^{1/2}$ | | db
0.0 | 5.38 | 7.31
0.5 | 6.40 | 8.06
1.0 | 10.94 | 10.4
1.5 | 27.1 | 14.3
2.0 | 89 | 19.5
ENERGY SPECTRUM OF THE INPUT NOISE: $e^{-\omega^2/\omega_b^2}$
$b_0/\psi^{1/2}$ =
2.0
1.5
1.0
0.5
0.0
SPECTRUM OF THE INPUT
$W(\frac{\omega}{\omega_b})$
$\frac{\omega}{\omega_b}$

FIG. 7. Output spectra for broad band input noise (I) after extreme clipping in a linear rectifier. (The relative levels in this figure and in Figs. 9, 14, 17, are the levels expressed in the (same) arbitrary units, *after* division by the factors indicated.)

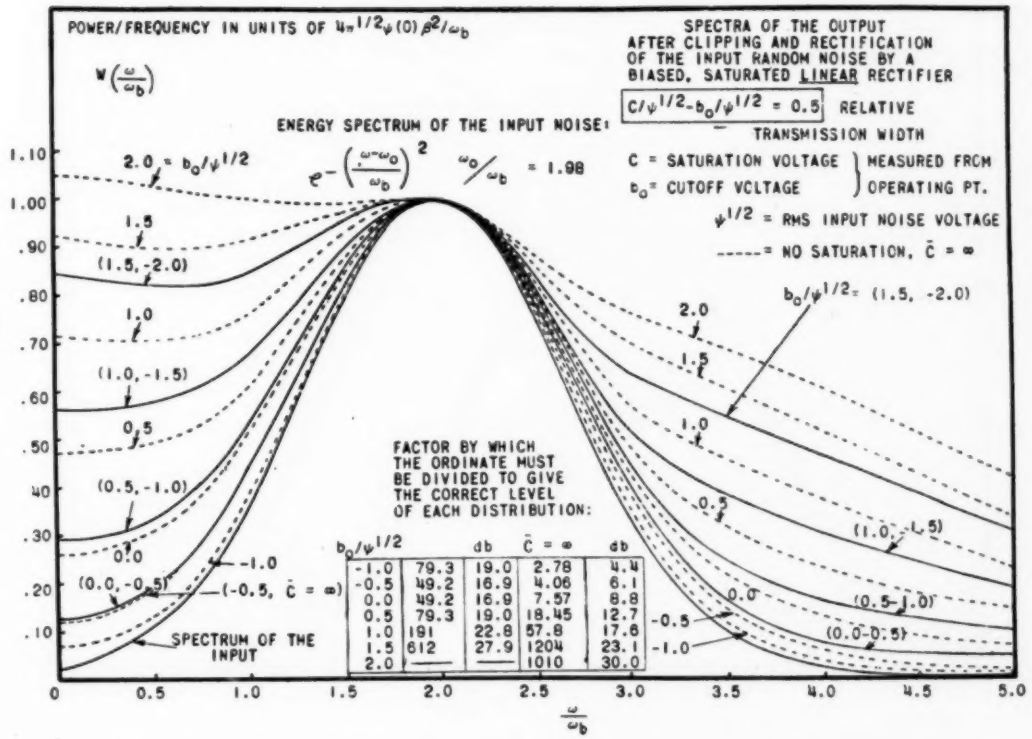

POWER/FREQUENCY IN UNITS OF $4\pi^{1/2}\psi(0)\beta^2/\omega_b$
SPECTRA OF THE OUTPUT
AFTER CLIPPING AND RECTIFICATION OF THE INPUT RANDOM NOISE BY A BIASED, SATURATED LINEAR RECTIFIER
$C/\psi^{1/2} - b_0/\psi^{1/2} = 0.5$ RELATIVE TRANSMISSION WIDTH
C = SATURATION VOLTAGE } MEASURED FROM
b_0 = CUTOFF VOLTAGE } OPERATING PT.
$\psi^{1/2}$ = RMS INPUT NOISE VOLTAGE
----- = NO SATURATION, $\bar{C} = \infty$
ENERGY SPECTRUM OF THE INPUT NOISE: $e^{-\left(\frac{\omega-\omega_0}{\omega_b}\right)^2}$ $\omega_0/\omega_b = 1.98$
$W(\frac{\omega}{\omega_b})$
$b_0/\psi^{1/2}$ = (1.5, -2.0)
2.0 = $b_0/\psi^{1/2}$
1.5
(1.5, -2.0)
1.0
(1.0, -1.5)
0.5
(0.5, -1.0)
0.0
(0.0, -0.5)
-1.0
(-0.5, $\bar{C} = \infty$)
SPECTRUM OF THE INPUT
FACTOR BY WHICH THE ORDINATE MUST BE DIVIDED TO GIVE THE CORRECT LEVEL OF EACH DISTRIBUTION:
$b_0/\psi^{1/2}$ | | db | $\bar{C} = \infty$ | db
-1.0 | 79.3 | 19.0 | 2.78 | 4.4
-0.5 | 49.2 | 16.9 | 4.06 | 6.1
0.0 | 49.2 | 16.9 | 7.57 | 8.8
0.5 | 79.3 | 19.0 | 18.45 | 12.7
1.0 | 191 | 22.8 | 57.8 | 17.6
1.5 | 612 | 27.9 | 1204 | 23.1
2.0 | — | — | 1010 | 30.0
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
(1.0, -1.5)
(0.5, -1.0)
(0.0, -0.5)
$\frac{\omega}{\omega_b}$

FIG. 8. Output spectra for semi-broad band noise (II), saturation and no saturation.

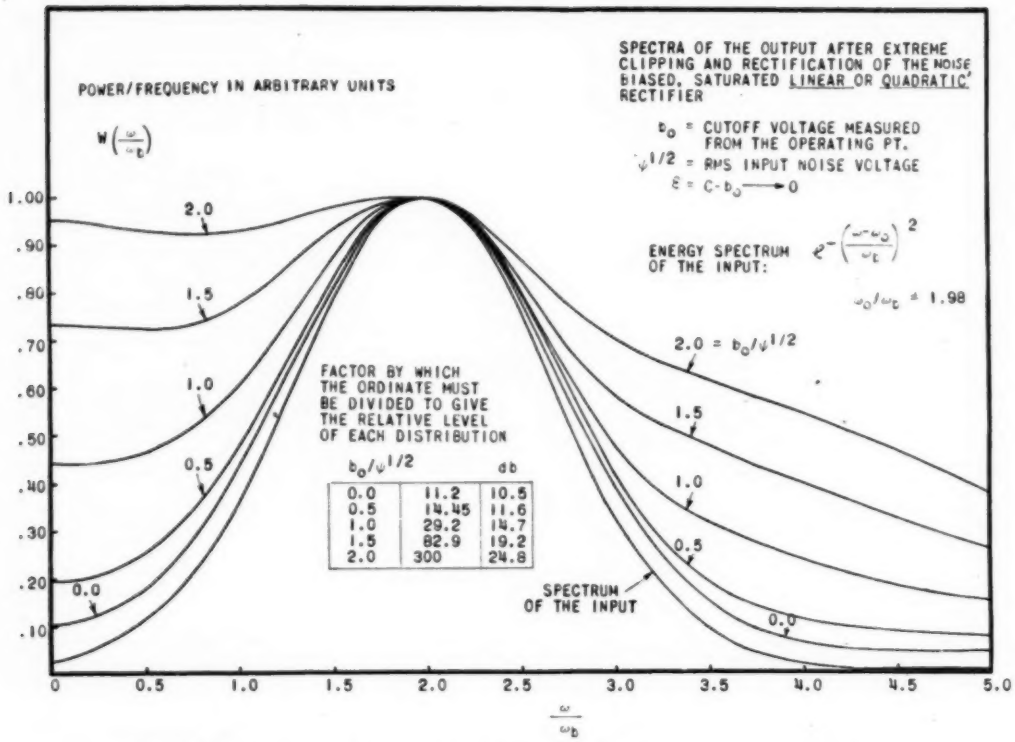


FIG. 9. Output spectra for semi-broad band noise (II) after extreme clipping.

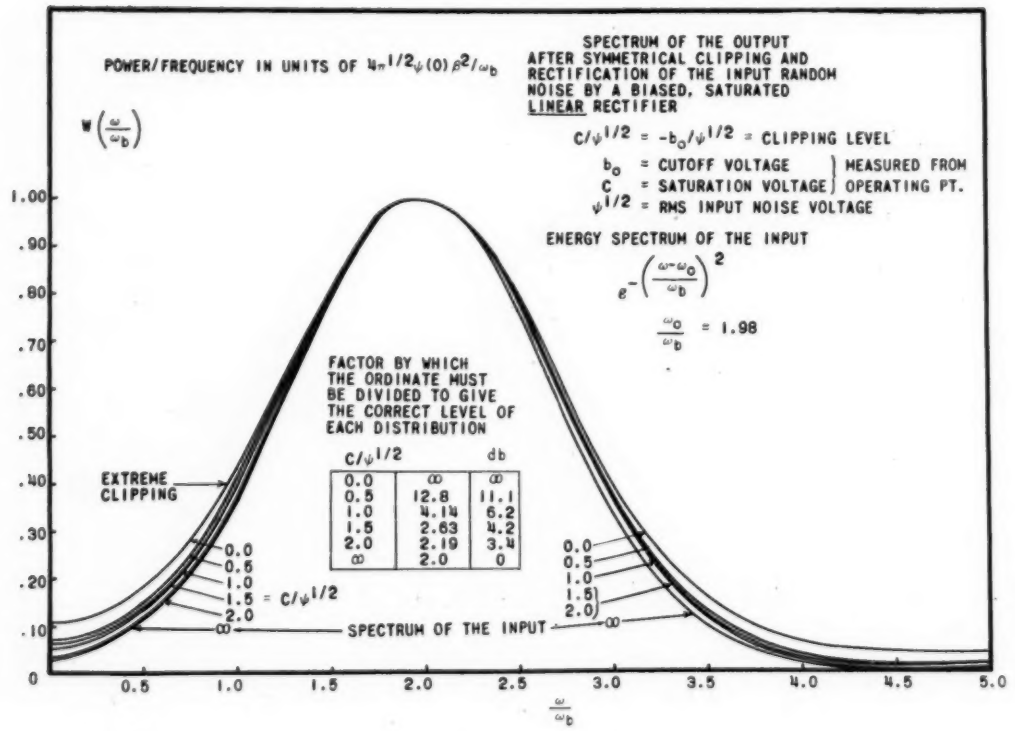


FIG. 10. Output spectra for semi-broad band noise (II) after symmetrical clipping.

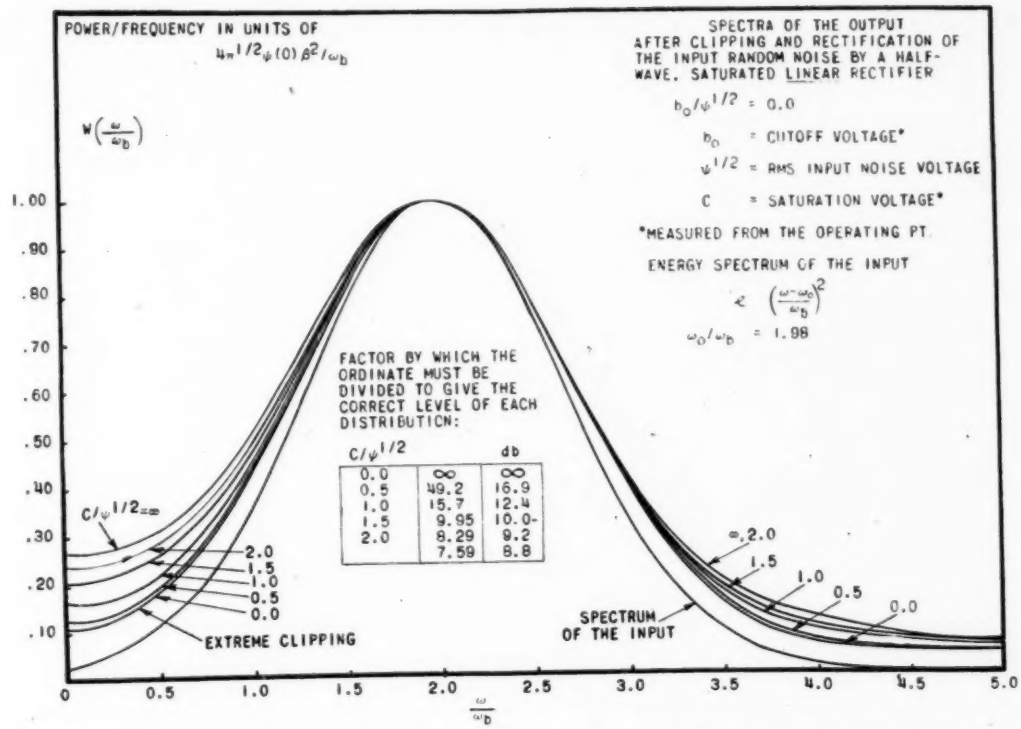

POWER/FREQUENCY IN UNITS OF
$4\pi^{1/2}\psi(0)\beta^2/\omega_b$
$W\left(\frac{\omega}{\omega_b}\right)$
SPECTRA OF THE OUTPUT AFTER CLIPPING AND RECTIFICATION OF THE INPUT RANDOM NOISE BY A HALF-WAVE, SATURATED LINEAR RECTIFIER
$b_0/\psi^{1/2} = 0.0$
b_0 = CUTOFF VOLTAGE*
$\psi^{1/2}$ = RMS INPUT NOISE VOLTAGE
C = SATURATION VOLTAGE*
*MEASURED FROM THE OPERATING PT.
ENERGY SPECTRUM OF THE INPUT
$\propto \left(\frac{\omega-\omega_0}{\omega_b}\right)^2$
$\omega_0/\omega_b = 1.98$
FACTOR BY WHICH THE ORDINATE MUST BE DIVIDED TO GIVE THE CORRECT LEVEL OF EACH DISTRIBUTION:
$C/\psi^{1/2}$ | | db
0.0 | ∞ | ∞
0.5 | 49.2 | 16.9
1.0 | 15.7 | 12.4
1.5 | 9.95 | 10.0
2.0 | 8.29 | 9.2
| 7.59 | 8.8
$C/\psi^{1/2}=\infty$
2.0
1.5
1.0
0.5
0.0
EXTREME CLIPPING
SPECTRUM OF THE INPUT
$\frac{\omega}{\omega_b}$

FIG. 11. Output spectra for semi-broad band noise (II) after half-wave rectification, with and without saturation.

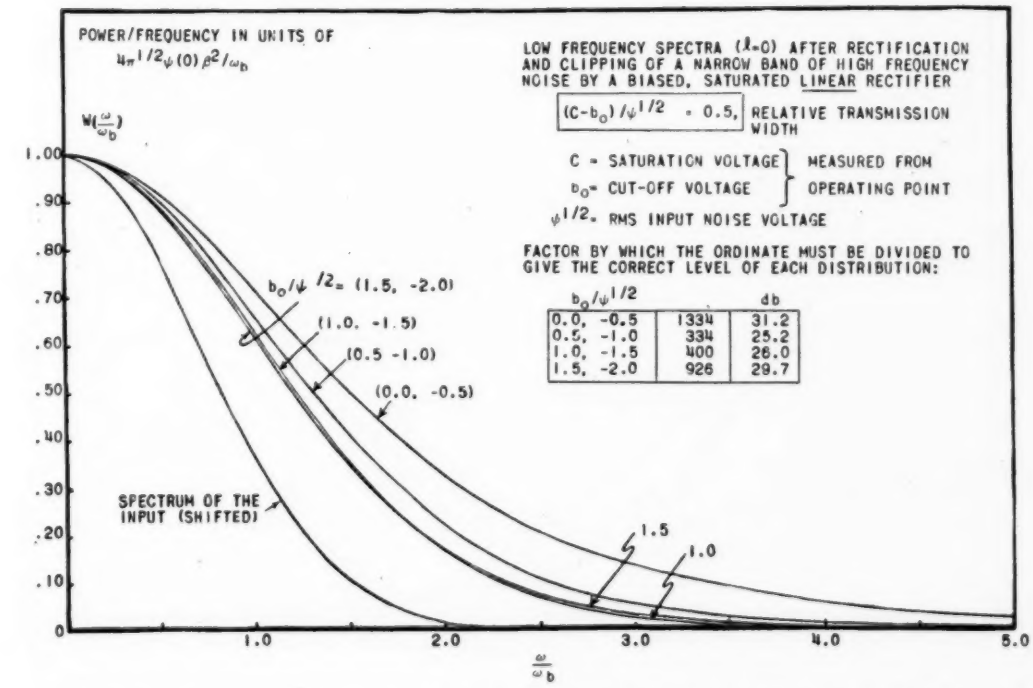

POWER/FREQUENCY IN UNITS OF
$4\pi^{1/2}\psi(0)\beta^2/\omega_b$
$W\left(\frac{\omega}{\omega_b}\right)$
LOW FREQUENCY SPECTRA (l=0) AFTER RECTIFICATION AND CLIPPING OF A NARROW BAND OF HIGH FREQUENCY NOISE BY A BIASED, SATURATED LINEAR RECTIFIER
$(C-b_0)/\psi^{1/2} = 0.5$, RELATIVE TRANSMISSION WIDTH
C = SATURATION VOLTAGE } MEASURED FROM
b_0 = CUT-OFF VOLTAGE } OPERATING POINT
$\psi^{1/2}$ = RMS INPUT NOISE VOLTAGE
FACTOR BY WHICH THE ORDINATE MUST BE DIVIDED TO GIVE THE CORRECT LEVEL OF EACH DISTRIBUTION:
$b_0/\psi^{1/2}$ | | db
0.0, -0.5 | 1334 | 31.2
0.5, -1.0 | 334 | 25.2
1.0, -1.5 | 400 | 26.0
1.5, -2.0 | 926 | 29.7
$b_0/\psi^{1/2}$ = (1.5, -2.0)
(1.0, -1.5)
(0.5 -1.0)
(0.0, -0.5)
SPECTRUM OF THE INPUT (SHIFTED)
1.5
1.0
$\frac{\omega}{\omega_b}$

FIG. 12. Low frequency output spectra (l=0), for narrow band input noise (III) with saturation.

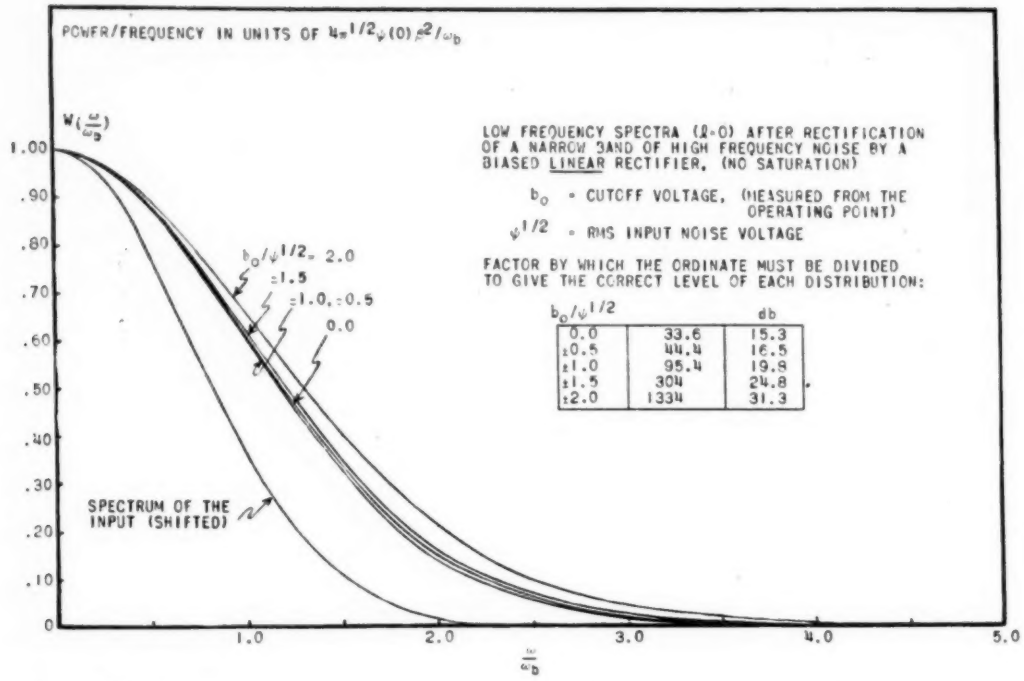


FIG. 13. Low frequency output spectra ($l=0$), for narrow band input noise (III) without saturation.

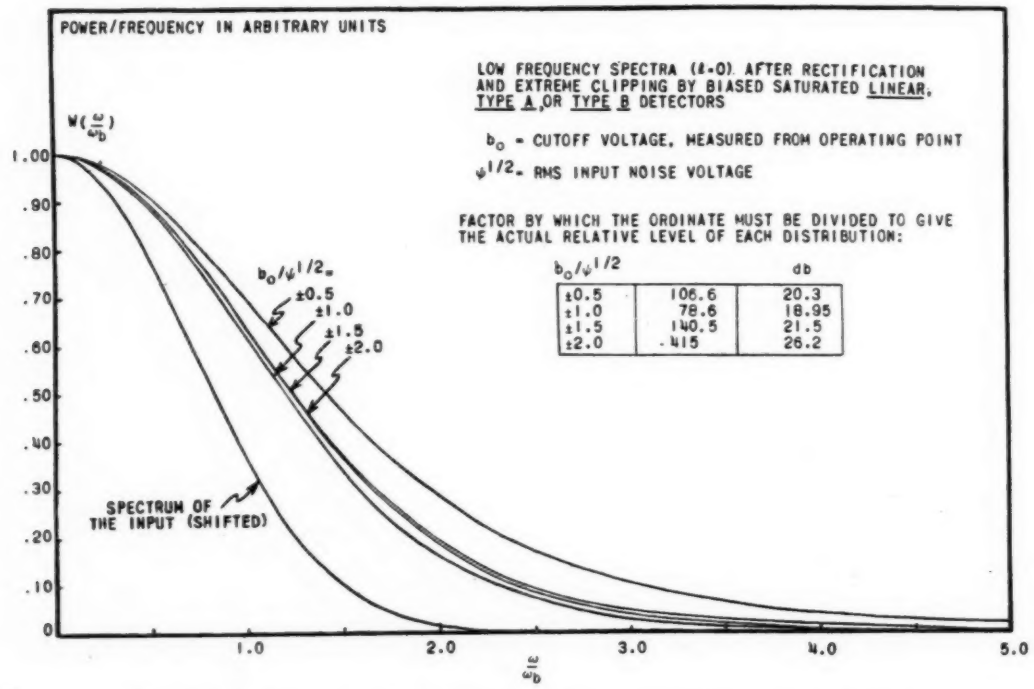


FIG. 14. Low frequency output spectra ($l=0$), for narrow band input noise (III) after extreme clipping.

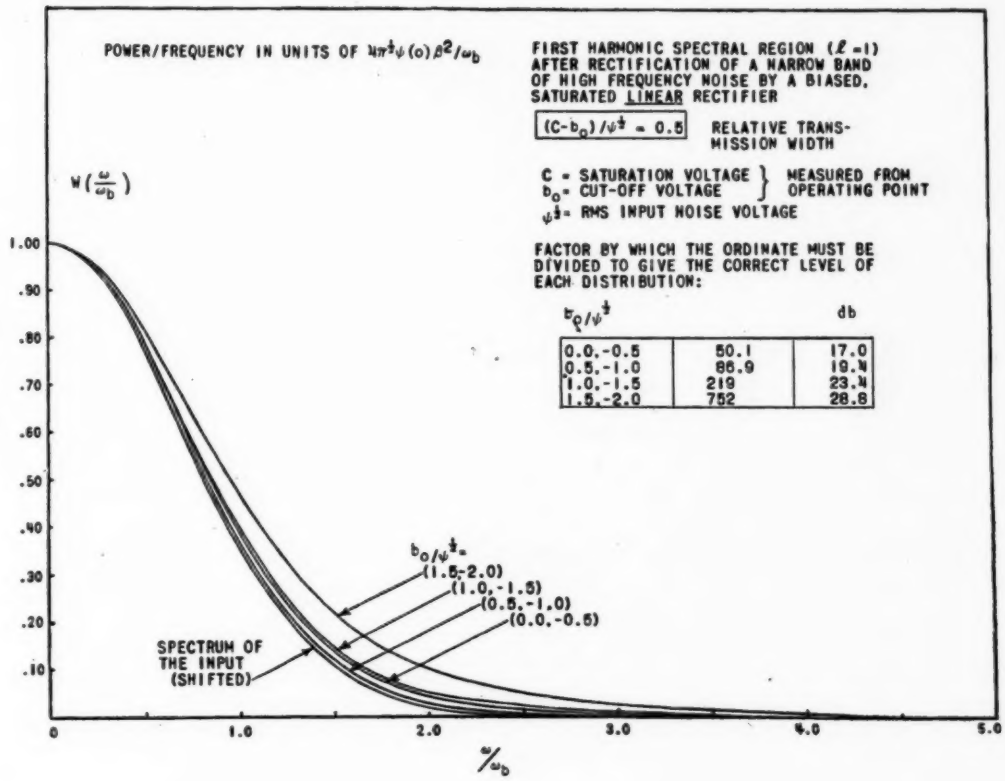


FIG. 15. High frequency output spectra ($l=1$), for narrow band input noise (III) with saturation.

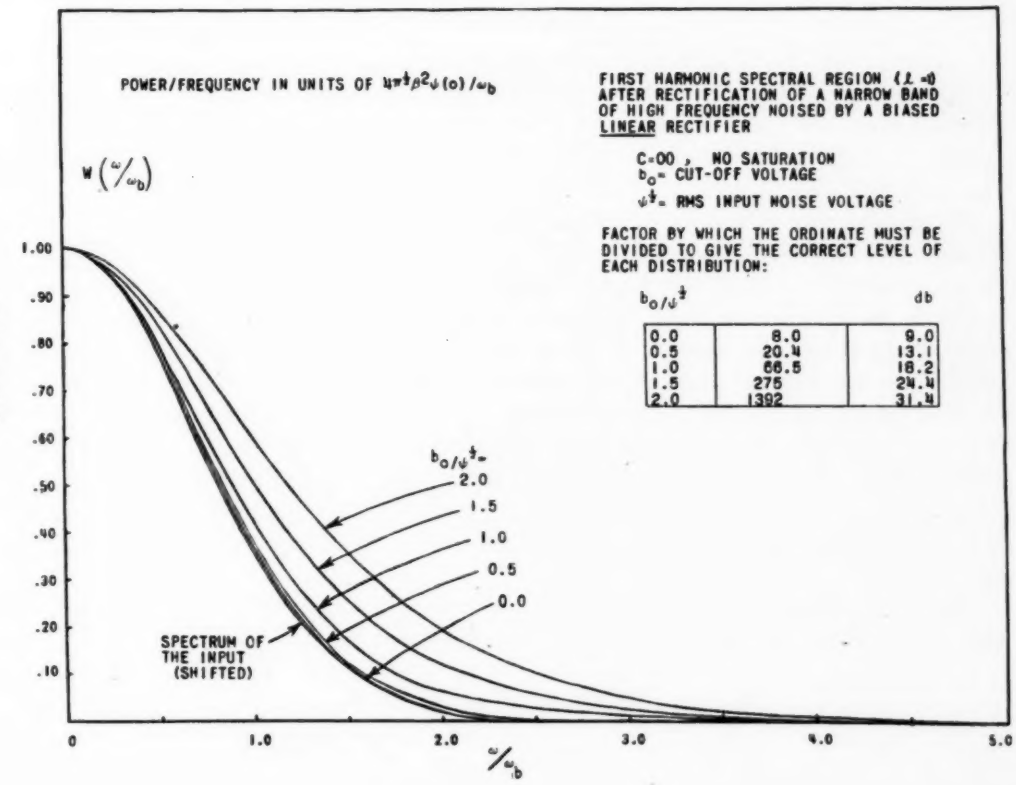


FIG. 16. High frequency output spectra ($l=1$), for narrow band input noise (III), no saturation.

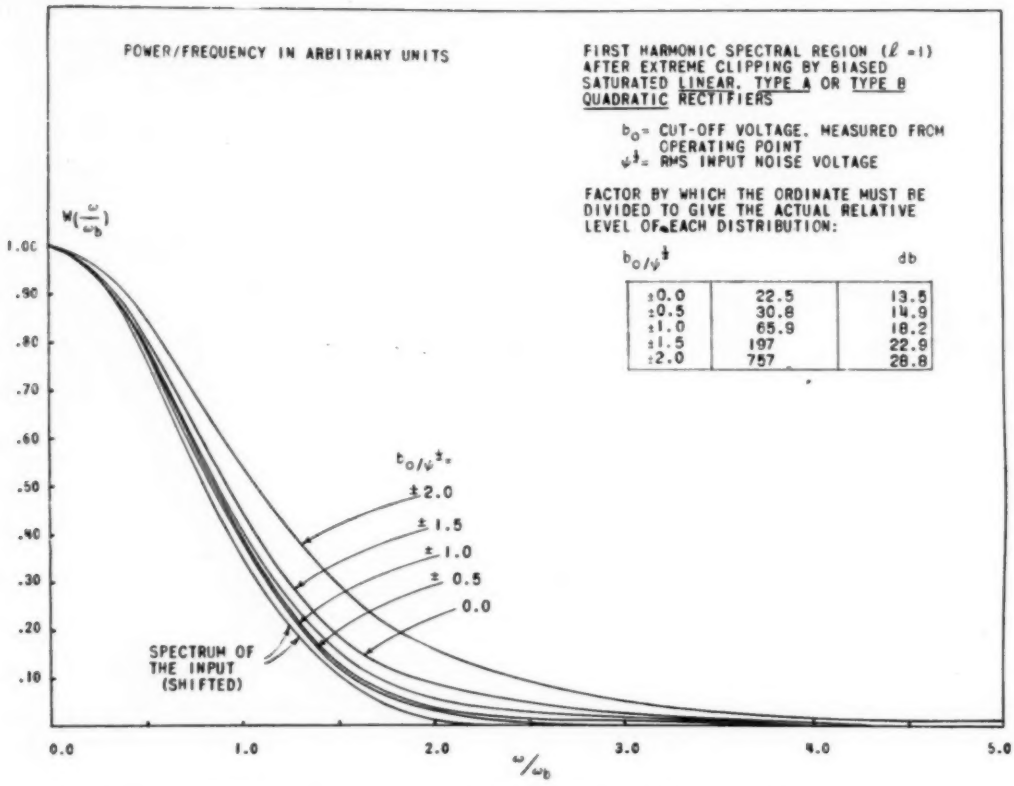


FIG. 17. High frequency output spectra ($l=1$), for narrow band input noise (III), extreme clipping.

A number of interesting special cases occur frequently in practice. The first is that of no saturation, a condition such that the instantaneous amplitudes exceed the critical value at which saturation occurs a negligible percentage of the time. Analytically $\bar{C}$ is then infinite, and Eqs. (4-3)–(4-6) are modified accordingly. Figures 6, 8, 13, 16 illustrate spectra in this instance; the dotted curves in Fig. 8 apply when $\bar{C}\to\infty$.

The second case is that of half-wave rectification where all input voltages V less than 0 are not passed, and all those for which $V>0$ are transmitted, *vide* Fig. 2. Then b_0 is required to be zero and the $h_{0,n}$ are appropriately simplified. Sample spectra are shown in Fig. 11.

Finally we have the case in which the input wave is clipped equally on the top and bottom. A study of the condition where $b_0=-\bar{C}$, as well as of the more general cases ($b_0<C$) should give some idea of the effect of saturation and cut-off on the output of an ordinary amplifier, as in practice Class A operation is often "contaminated" by some (nearly symmetrical) clipping of the larger amplitudes, due to the leveling off of the dynamic response. In all instances where the input is random noise we have departures from pure Class A performance. Here we observe that $h_{0,n}$, $n=2, 4, 6, \cdots$ is zero, since then $\phi^{(n-2)}$ is an even function. In other words, of the components that contribute to the spectrum none of the even ones are generated. Figure 10 indicates some typical spectra.

Among the results should be included the correlation function for linear half-wave rectification when there is no saturation. Equation (3-4) then reduces to

$$R(t)_{b_0=0}=\frac{\beta^2\psi}{2\pi}\left\{1+\pi r/2+\sum_{n=2,4,\cdots}^{\infty}\left[\frac{(n-2)!}{2^{(n-2)/2}\left(\frac{n-2}{2}\right)!}\right]^2 r^n/n!\right\}, \tag{4-12}$$

which may be expressed in closed form:

$$R(t)_{b_0=0}=\frac{\beta^2\psi}{2\pi}\{(1-r^2)^{\frac{1}{2}}+r(\pi/2+\sin^{-1}r)\}, \tag{4-13}$$

a result equivalent to Eq. (4.7-5) of Rice's paper.

It is instructive to examine a number of limiting cases. If the cut-off and saturation voltages are equal there should be no output, for obvious physical reasons. Setting $b_0\equiv C$ in (4-3)–(4-6) or more directly in (4-1) shows that this is indeed the case. Then, too, we may let C approach infinity in (4-11), when $b_0=-C$. This corresponds to biasing so positively, i.e., $b_0\ll 0$, that the entire disturbance is passed with negligible distortion. From (4-3) we see that $W_{d.c.}=C^2\beta^2$; therefore, the power associated with the continuum is $W_c=R(0)-C^2\beta^2=\beta^2\psi$, as $\bar{C}\to\infty$, in other words, the input wave is completely transmitted. The spectrum is also unchanged, a fact easily verified from (4-3)–(4-6), since only $h_{0,1}$ does not vanish when $\bar{C}\to\infty$, with the exception of the d.c. contribution, whose amplitude is governed by $h_{0,0}$. Of course (in theory), infinite power must be supplied by the bias source to shift the response so that (essentially) infinitely negative noise peaks will be passed. That we obtain an infinite total power as $\bar{C}\to\infty$ is clear from (4-11). Still another test may be made if we allow the bias to become greatly negative, i.e., $b_0\gg 0$, with the condition that the saturation level $\bar{C}$ always exceed the cut-off voltage, i.e. $\bar{C}>\bar{b}_0(\gg 0)$. Physically we expect a vanishing output, for only the very infrequent, large noise amplitudes will be passed, and as the bias is made more and more unfavorable, the probability of the occurrence of these becomes negligible. Examination of (4-3)–(4-6) does indeed show that $h_{0,n}\to 0$ as $\bar{C}$, $\bar{b}_0\to\infty$, so that under such conditions there is no output.

V. THE TYPE-B QUADRATIC RECTIFIER

We may determine the Fourier transform $f(iz)_B$ of the Type-B dynamic path with the help of (1-2), by expressing $f(iz)$ as a polynomial in V, since from (1-2), $I(t)$ is a polynomial function of $V(t)$ over specified ranges of V. We start with a representation of the form

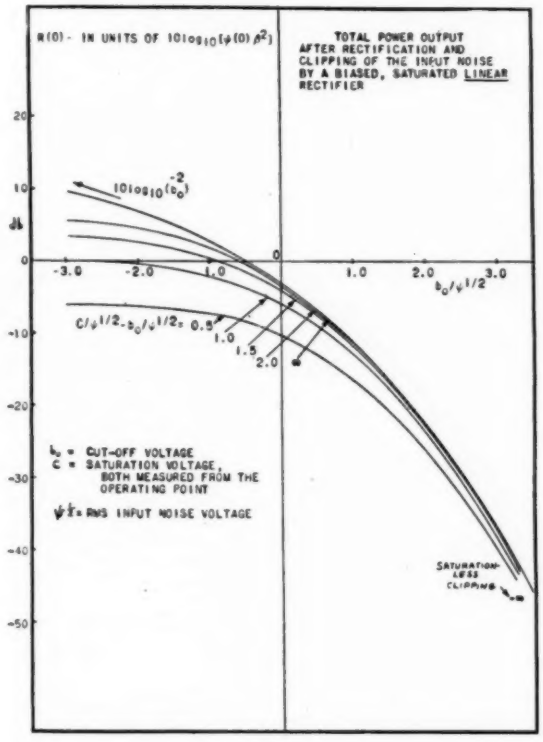

R(0) - IN UNITS OF $10\log_{10}[\psi(0)\beta^2]$
TOTAL POWER OUTPUT AFTER RECTIFICATION AND CLIPPING OF THE INPUT NOISE BY A BIASED, SATURATED LINEAR RECTIFIER
$10\log_{10}(b_0^{-2})$
$b_0/\psi^{1/2}$
$C/\psi^{1/2}-b_0/\psi^{1/2}=$ 0.5 1.0 1.5 2.0 ∞
b_0 = CUT-OFF VOLTAGE
C = SATURATION VOLTAGE, BOTH MEASURED FROM THE OPERATING POINT
$\psi^{\frac{1}{2}}$ = RMS INPUT NOISE VOLTAGE
SATURATION-LESS CLIPPING
-∞

FIG. 18. Total power output of a linear rectifier for input noise of types (I–III).

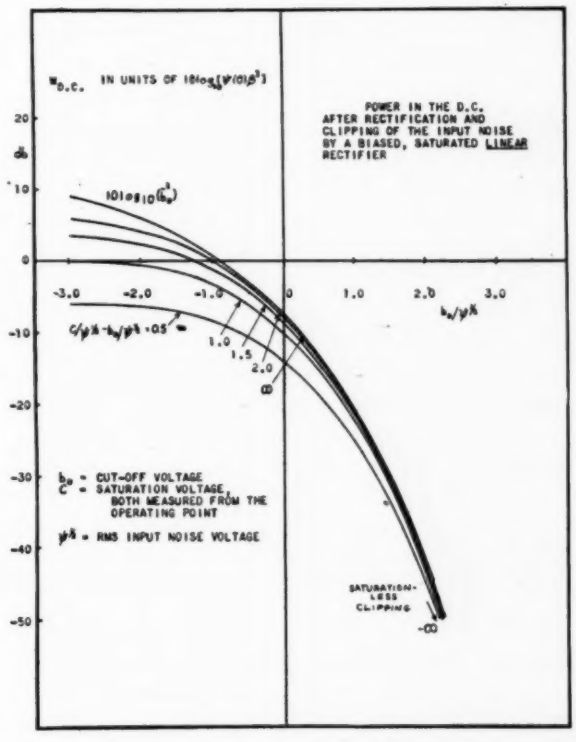

$W_{d.c.}$ IN UNITS OF $10\log_{10}[\psi(0)\beta^2]$
POWER IN THE D.C. AFTER RECTIFICATION AND CLIPPING OF THE INPUT NOISE BY A BIASED, SATURATED LINEAR RECTIFIER
$10\log_{10}(b_0^{-2})$
$b_0/\psi^{\frac{1}{2}}$
$C/\psi^{\frac{1}{2}}-b_0/\psi^{\frac{1}{2}}=$ 0.5 1.0 1.5 2.0 ∞
b_0 = CUT-OFF VOLTAGE
C = SATURATION VOLTAGE, BOTH MEASURED FROM THE OPERATING POINT
$\psi^{\frac{1}{2}}$ = RMS INPUT NOISE VOLTAGE
SATURATION-LESS CLIPPING
-∞

FIG. 19. Power in the d.c. output of a linear rectifier for input noise of types (I–III).

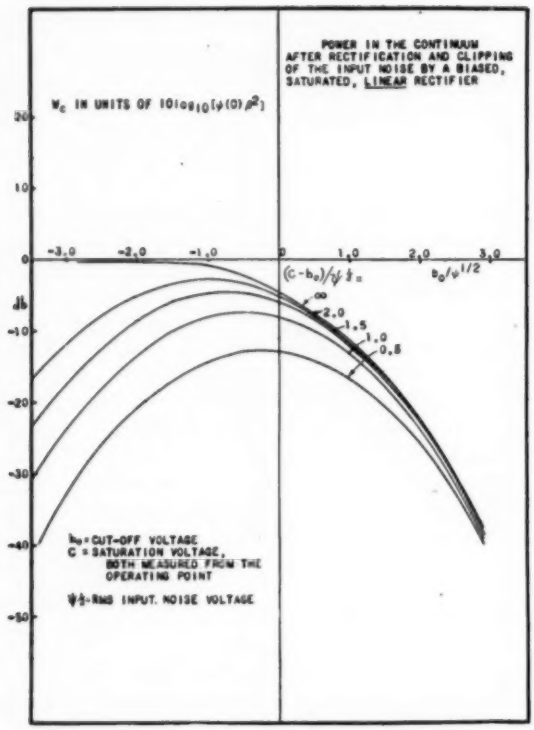

POWER IN THE CONTINUUM AFTER RECTIFICATION AND CLIPPING OF THE INPUT NOISE BY A BIASED, SATURATED, LINEAR RECTIFIER
W_c IN UNITS OF $10\log_{10}[\psi(0)\beta^2]$
$(C-b_0)/\psi^{\frac{1}{2}}=$
$b_0/\psi^{1/2}$
b_0 = CUT-OFF VOLTAGE
C = SATURATION VOLTAGE, BOTH MEASURED FROM THE OPERATING POINT
$\psi^{\frac{1}{2}}$ = RMS INPUT NOISE VOLTAGE

FIG. 20. Power in the continuous portion of the output of a linear rectifier, for input noise of types (I), (III).

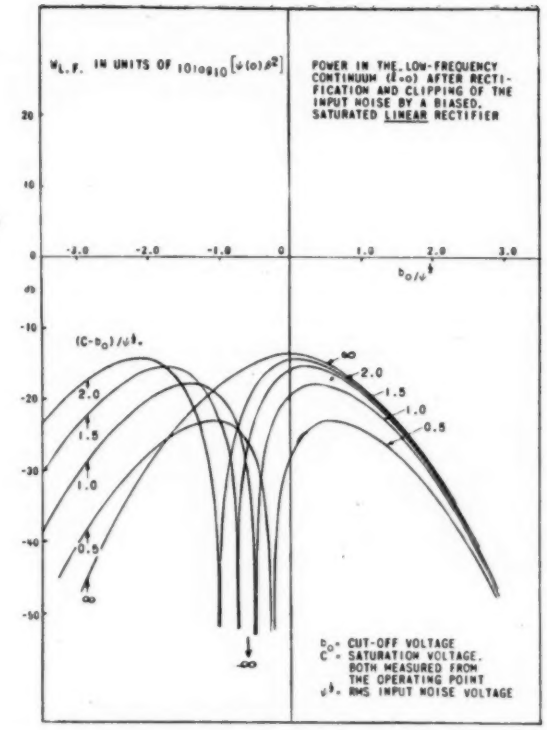

$W_{L.F.}$ IN UNITS OF $10\log_{10}[\psi(o)\beta^2]$
POWER IN THE LOW-FREQUENCY CONTINUUM (f=o) AFTER RECTIFICATION AND CLIPPING OF THE INPUT NOISE BY A BIASED, SATURATED LINEAR RECTIFIER
$(C-b_0)/\psi^{\frac{1}{2}}$
$b_0/\psi^{\frac{1}{2}}$
b_0 = CUT-OFF VOLTAGE
C = SATURATION VOLTAGE, BOTH MEASURED FROM THE OPERATING POINT
$\psi^{\frac{1}{2}}$ = RMS INPUT NOISE VOLTAGE

FIG. 21. Power output in the low frequency ($f=0$) continuum for type III noise entering a linear rectifier.

$$I(t)=\frac{\lambda^2}{2\pi}\int_C\left\{\frac{2e^{i(V-b_0)z}+2Ae^{i(V-2C+b_0)}+2Be^{i(V-C)z}}{(iz)^3}+\frac{De^{i(V-C)z}}{(iz)^2}+\frac{Ee^{i(V-C)z}}{(iz)}\right\}dz, \quad C\geq b_0. \tag{5-1}$$

To determine A, B, D, and E, we evaluate (5-1) for the various regions involved and match these results to the desired solution for each region. For example, when $V>2C-b_0$, $I(t)=2\lambda^2(C-b_0)^2$, and so

$$(V-b_0)^2+A(V-2C+b_0)^2+B(V-C)^2+D(V-C)+E=2(C-b_0)^2, \tag{5-2}$$

from which we see that $A=1$, $B=-2$, $D=E=0$. The values of A, B, D, and E so obtained are easily seen to be compatible with the other regions $V<2C-b_0$. In this manner we find for the Type-B detector that $f(iz)$ is given by

$$f(iz)_B=2\lambda^2\{\exp(-ib_0z)+\exp[-i(2C-b_0)z]-2\exp(-iCz)\}/(iz)^2, \quad C\geq b_0. \tag{5-3}$$

Then (3-3) may be written

$$h_{0,n}=\frac{i\lambda^2}{\pi}\int_C\exp(-\psi z^2/2)z^{n-3}\{\exp(-ib_0z)+\exp[-i(2C-b_0)z]-2\exp(-iCz)\}dz, \quad n\geqq 0. \tag{5-4}$$

Equation (5-4) may be integrated with the help of the Appendix, to give

$$h_{0,0}=-2\lambda^2\psi\{(\bar{b}_0^2+1)(\Theta_{b_0}-1)+2\bar{b}_0\phi_{b_0}+[(2C-b_0)^2+1][\Theta_{2c-b_0}-1]+2(2\bar{C}-\bar{b}_0)\phi_{2C-b_0}-2(\bar{C}^2+1)(\Theta_c-1)+4\phi_c\}, \tag{5-5}$$

$$h_{0,1}=-i\lambda^2\psi^{\frac{1}{2}}\{\bar{b}_0(\Theta_{b_0}-1)+2\phi_{b_0}+(2\bar{C}-\bar{b}_0)(\Theta_{2c-b_0}-1)+2\phi_{2C-b_0}-2\bar{C}(\Theta_c-1)-4\phi_c\}, \tag{5-6}$$

$$h_{0,2}=\lambda^2\{\Theta_{b_0}+\Theta_{2c-b_0}-2\Theta_c\}, \tag{5-7}$$

$$h_{0,n}=2i\lambda^2(-1)^{(n-3)/2}\psi^{(2-n)/2}\{\phi_{b_0}{}^{(n-3)}+\phi_{2C-b_0}{}^{(n-3)}-2\phi_c{}^{(n-3)}\}, \quad n=3,5,7,\cdots, \tag{5-8}$$

$$h_{0,n}=2\lambda^2(-1)^{(n-2)/2}\psi^{(2-n)/2}\{\phi_{b_0}{}^{(n-3)}+\phi_{2C-b_0}{}^{(n-3)}-2\phi_c{}^{(n-3)}\}, \quad n=4,6,8,\cdots. \tag{5-9}$$

We have made the additional abbreviations

$$\phi_{2C-b_0}^{(j)} \equiv \phi^{(j)}[(2C-b_0)/\psi^{\frac{1}{2}}]; \quad \Theta_{2c-b_0} \equiv \Theta[(2C-b_0)/(2\psi)^{\frac{1}{2}}]. \tag{5-10}$$

The total output power of the Type-B quadratic device may be found from (1-2) and (3-8). We write then

$$W_\tau = R(0) = \frac{\lambda^4\psi^2}{(2\pi)^{\frac{1}{2}}}\left\{\int_{\bar{b}_0}^{C/\psi^{\frac{1}{2}}} (x-\bar{b}_0)^4 \exp(-x^2/2)dx + \int_{\bar{c}}^{2\bar{C}-b_0/\psi^{\frac{1}{2}}} [2(\bar{C}-\bar{b}_0)^2-(x-2\bar{C}+b_0)^2]^2 \exp(-x^2/2)dx \right.$$
$$\left. +4(\bar{C}-\bar{b}_0)^4\int_{2\bar{c}-\bar{b}_0}^{\infty} \exp(-x^2/2)dx\right\}. \tag{5-11}$$

This we integrate to obtain finally

$$W_\tau = \frac{\lambda^4\psi^2}{2}\{P_1\Theta_{2c-b_0}-P_2\Theta_c-P_3\Theta_{b_0}+2P_4\phi_{2C-b_0}+2P_5\phi_c-2P_6\phi_{b_0}+4(\bar{C}-b_0)^4\}. \tag{5-12}$$

where

$$P_1 = 16\bar{C}^3\bar{b}_0-28\bar{C}^2\bar{b}_0^2+16\bar{C}\bar{b}_0^3-3\bar{b}_0^4+20\bar{C}^2-16\bar{C}+2\bar{b}_0^2+3, \tag{5-13a}$$

$$P_2 = 4\bar{C}^4-4\bar{C}^2\bar{b}_0^2-4\bar{b}_0^2+20\bar{C}^2-16\bar{C}\bar{b}_0, \tag{5-13b}$$

$$P_3 = \bar{b}_0^4+6\bar{b}_0^2+3, \tag{5-13c}$$

$$P_4 = 3\bar{b}_0^3-5\bar{b}_0+8\bar{b}_0\bar{C}^2-10\bar{C}\bar{b}_0^2+10\bar{C}. \tag{5-13d}$$

$$P_5 = 4(4\bar{b}_0-4\bar{C}+\bar{b}_0^2\bar{C}-\bar{C}^3), \tag{5-13e}$$

$$P_6 = \bar{b}_0^3+5\bar{b}_0. \tag{5-13f}$$

The power in the continuum follows when $h_{0,0}^2$, Eq. (5-5), is subtracted from (5-12).

It is interesting to observe for symmetrical clipping, $C=0$, $b_0<0$, that the Type-B rectifier yields no output in the even harmonic zones $l=0, 2, 4, \cdots$, when the input noize is a narrow band. Only odd-order products contribute on the average, as is true under similar circumstances in the case of the linear detector, when $b_0=-C$.

As before, when the cases of half-wave rectification ($b_0=0$), or symmetrical clipping (which does not obtain for Type A responses), is discussed, considerable simplification results in the expression for the output power and spectra. An important example is the half-wave, unsaturated quadratic detector. For this device we obtain the output correlation function

$$R(t)_{b_0=0} = \frac{\lambda^4\psi^2}{2\pi}\left\{\frac{\pi}{2}+4r+\pi r^2+4\sum_{n=3,5,\cdots}^{\infty}\left[\frac{(n-3)!}{2^{(n-3)/2}\left(\frac{n-3}{2}\right)!}\right]^2 r^n/n!\right\}, \tag{5-14}$$

which becomes, in closed form,

$$R(t)_{b_0=0} = \frac{\lambda^4\psi^2}{2\pi}\left\{\left(\frac{\pi}{2}+\sin^{-1}r\right)(1+2r^2)+3r(1-r^2)^{\frac{1}{2}}\right\}. \tag{5-15}$$

VI. EXTREME CLIPPING

The limiting case of extreme or "superclipping" at the "top" and "bottom" of the input is instructive as it offers an answer to such questions as whether unrestricted clipping results in an indefinitely great spectral spread, and whether the spectra so obtained appear with greatly reduced levels. Some cases of symmetrical clipping, including the limiting extreme of complete "slicing" of the input wave, have been considered by J. H. Van Vleck. We are now able to extend these earlier results to the general case of superclipping at any arbitrary cut-off level, i.e., for arbitrary values of b_0. Analytically, extreme clipping means that the saturation level C approaches the cut-off b_0 so

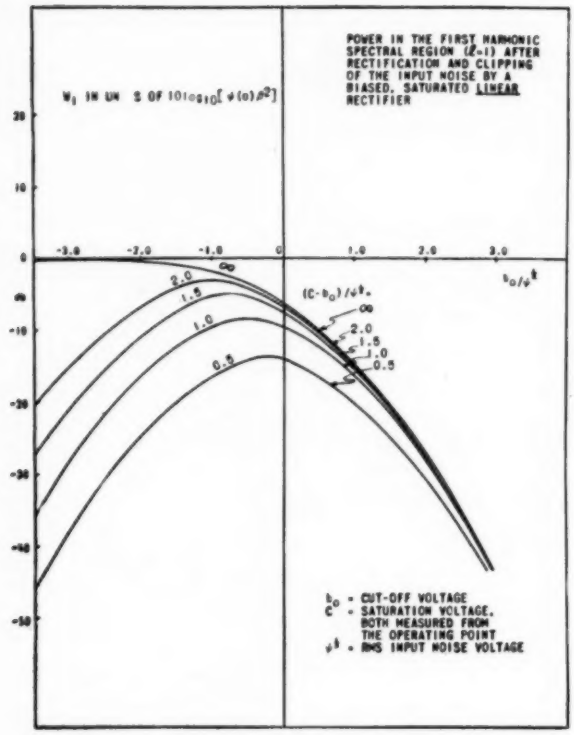

POWER IN THE FIRST HARMONIC SPECTRAL REGION (l=1) AFTER RECTIFICATION AND CLIPPING OF THE INPUT NOISE BY A BIASED, SATURATED LINEAR RECTIFIER
W₁ IN DB OF 10 log₁₀[ψ(0)β²]
b₀/ψ^½
(C-b₀)/ψ^½
∞
2.0
1.5
1.0
0.5
b₀ : CUT-OFF VOLTAGE
C : SATURATION VOLTAGE, BOTH MEASURED FROM THE OPERATING POINT
ψ^½ : RMS INPUT NOISE VOLTAGE

FIG. 22. Power output in the high frequency ($l=1$) continuum for type III noise entering a linear rectifier.

that the voltage difference between the two levels is infinitesimal. Then for all types of superclipping the following conditions must hold:

$$C-b_0=\epsilon, \text{ or } \bar{C}-\bar{b}_0=\tilde{\epsilon};\ C>b_0,\ 0<\epsilon\ll 1. \quad (6\text{-}1)$$

We may obtain the $h_{0,n}$'s from (4-3)–(4-6) by applying Taylor's theorem. Let $g(\bar{C})$ be the function whose value at $\bar{C}=\bar{b}_0+\tilde{\epsilon}$ we wish to determine. Expansion about the point $\bar{b}_0$ gives

$$g(\bar{C})=g(\bar{b}_0)+\tilde{\epsilon}g'(\bar{b}_0)+\tilde{\epsilon}^2g''(\bar{b}_0)/2+\cdots. \quad (6\text{-}2)$$

It is readily verified for the linear detector that only terms in $\tilde{\epsilon}$ are significant; the higher powers of $\tilde{\epsilon}$ are negligible and the zeroth power ($\tilde{\epsilon}^0=1$) vanishes in $h_{0,n}$. Equations (4-3)–(4-6) are accordingly modified to

$$h_{0,0}\doteq-\beta\tilde{\epsilon}\psi^{\frac{1}{2}}(\Theta_{b_0}-1)/2, \quad (6\text{-}3)$$

$$h_{0,1}\doteq-\beta i\tilde{\epsilon}\phi_{b_0}, \quad (6\text{-}4)$$

$$h_{0,n}\doteq\beta(-1)^{(n+2)/2}\psi^{(1-n)/2}\tilde{\epsilon}\phi_{b_0}{}^{(n-1)},\quad n=2,4,6,\cdots, \quad (6\text{-}5)$$

$$h_{0,n}\doteq\beta i(-1)^{(n+3)/2}\psi^{(1-n)/2}\tilde{\epsilon}\phi_{b_0}{}^{(n-1)},\quad n=3,5,7,\cdots. \quad (6\text{-}6)$$

Spectra, examples of which are shown in Figs. 7, 9, 14, 17, may be obtained from (3-1) and (3-2), with the help of (6-3)–(6-6). It can also be demonstrated for the quadratic rectifier, discussed briefly in Section V, that $h_{0,0}\sim(\Theta_{b_0}-1)/2$ and $h_{0,n}\sim\phi_{b_0}{}^{(n-1)}$, in line with the results above, the proportionality constant $\beta\tilde{\epsilon}$ being replaced by $2\lambda^2\tilde{\epsilon}^2$, indicating that the *shape* of the output spectrum is independent of the dynamic i_p-e_g path if $\tilde{\epsilon}$ is taken sufficiently small, a reasonable conclusion, based on the observation that the output is now essentially a square wave with but two possible amplitudes, determined by the limit of the transmission width and not by the dynamic path between those limits.

The correlation function of the output in cases of extreme clipping is for *any* input distribution,

$$R(t)\doteq\tilde{\epsilon}^2\psi\beta^2\left\{\left(\frac{1-\Theta_{b_0}}{2}\right)^2+\sum_{n=1}^{\infty}[\phi_{b_0}{}^{(n-1)}]^2r(t)^n/n!\right\}. \quad (6\text{-}7)$$

The total power may be obtained as before by setting $t=0$ in (6-7), or by applying (4-10). We have

$$W_\tau=R(0)\doteq\beta^2\tilde{\epsilon}^2(1-\Theta_{b_0})/2. \quad (6\text{-}8)$$

The power in the continuum is therefore

$$W_c=R(0)-h_{0,0}{}^2\doteq\beta^2\tilde{\epsilon}^2(1-\Theta_{b_0}{}^2)/4. \quad (6\text{-}9)$$

If we specialize our results to symmetrical clipping we need only multiply (6-7)–(6-9) by 4. This follows because the effective amplitude of the disturbance is doubled, and hence the power is increased fourfold. When the incoming wave undergoes half-wave clipping and rectification, b_0 is zero, and Eq. (6-7) reduces to

$$R(t)_{b_0=0}\doteq\tilde{\epsilon}^2\beta^2\left\{\frac{1}{4}+\frac{1}{2\pi}\sum_{n=1,3,\cdots}^{\infty}\times\left\{(n-1)!\Big/2^{(n-1)/2}\left(\frac{n-1}{2}\right)!\right\}^2r(t)^n/n!\right\}, \quad (6\text{-}10)$$

which may be written as

$$R(t)_{b_0=0}=\tilde{\epsilon}^2\beta^2\left(\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}r\right)=\tilde{\epsilon}^2\beta^2\left\{\frac{1}{4}+\frac{r}{2\pi}\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};\tfrac{3}{2};r^2)\right\}, \quad (6\text{-}11)$$

a result similar to that obtained by Van Vleck. A rather interesting consequence of extreme clipping is that zero cut-off voltage ($b_0=0$), gives the greatest actual power in the continuum ($\bar{\epsilon}$ non-vanishing), while for $b_0\to\infty$, i.e., very negative bias, a greater *percentage* of the power goes into the continuum.

VII. RESULTS AND CONCLUSIONS

In the preceding sections we have outlined the analysis from which the mean power and spectral output of biased and saturated linear or quadratic rectifiers may be determined when the incoming wave is random noise. Since much of our work appears specifically in the figures, it is with them principally in mind that we discuss the results. The transmission width $C-b_0$, or $2(C-b_0)$, is treated as constant, which it is for resistive loads, so that as the amount of incoming noise is varied, the *relative* transmission width $(C-b_0)/\psi^{\frac{1}{2}}$, or $2(C-b_0)/\psi^{\frac{1}{2}}$, takes on different values. This corresponds physically to the operation of a given tube with various applied grid voltages and bias. The effects of clipping on the output spectra have been compared by normalizing the maximum spectral level of the output to unity. The normalization is accomplished after multiplication by the appropriate quantity, here donated by A. For example, in Fig. 8 when $(C-b_0)/\psi^{\frac{1}{2}}=0.5$ and $b_0/\psi^{\frac{1}{2}}=1.5$, $(o-2.0)$, A is seen to be 612 or its equivalent 27.9 db, where the db are determined from 10 $\log_{10} A$. To obtain the absolute spectral level we have only to divide the normalized value by A.

Extensive calculations and figures corresponding to Figs. 5-23 have also been completed for the Type A and Type B quadratic rectifiers. It has been found that the results in the latter cases are qualitatively similar to those for the linear rectifier, as may be seen in a few special instances, cf. Figs. 24 and 25, on comparison with Figs. 5 and 18. The general shape of the spectra is much the same, and the various output power distributions, analagous to Figs. 18–22 are similar. A comparison of absolute levels cannot in general be made, however, as it requires a specific knowledge of the tube constants β and γ, λ, which in turn are functions of the operating conditions. Because of general similarity of the results for the three types of detector, we have omitted the detailed analyses and the figures in the square-law cases, but these are available on request.

(a) Spectra

The first observation to be made from Figs. 5–17 is that clipping, whether at the "top" or "bottom" of the input wave, *always* results in a spectral spread.* This may be appreciated in a qualitative fashion from the following simple example. Consider first the spectrum of a sine wave. It will contain just one component, the original wave. Now clip the wave, at the top or bottom, or at both places. The signal is distorted and has "corners" where the clipping took place. When the disturbance is analyzed, we find a large number of new, high frequency terms; these are necessary to give the "squared-off" shape of the wave. Physically, a fraction of the input power is lost, since part of the signal is removed, and the rest of the input power is distributed among the higher harmonics generated by clipping, the precise redistribution depending on the wave's amplitude, the tube characteristic, etc. The above example is easily extended to noise with its infinite number of components.

The curves of Fig. 8 also show that the spectral spread is greatest when clipping occurs because of the tube cut-off alone and not on account of saturation as well. There are a few exceptions to this, which occur only for narrow-band input noise, as Figs. 13 and 14 show. Note that increasing the relative transmission width $(C-b_0)/\psi^{\frac{1}{2}}$ yields clipping which arises principally from cut-off, for a given value of b_0, since overloading effects then tend to disappear. When the rectified noise output is passed into amplifiers the level alone is altered and the character of the noise is preserved, *provided* the amplification is a purely linear process. In this connection we observe that filters, of which amplifiers are an example, discriminate on a frequency basis, and affect the instantaneous amplitude of the wave only indirectly as they respond or fail to respond to the various components. Filtering tends to

* That clipping always spreads the spectrum in the case of noise may be proved by summing the limiting forms of $h_{0,n}{}^2c_{0,n}(f)$ as $\bar{b}_0\to\infty$, and comparing the normalized sum with that for finite values of $\bar{b}_0$.

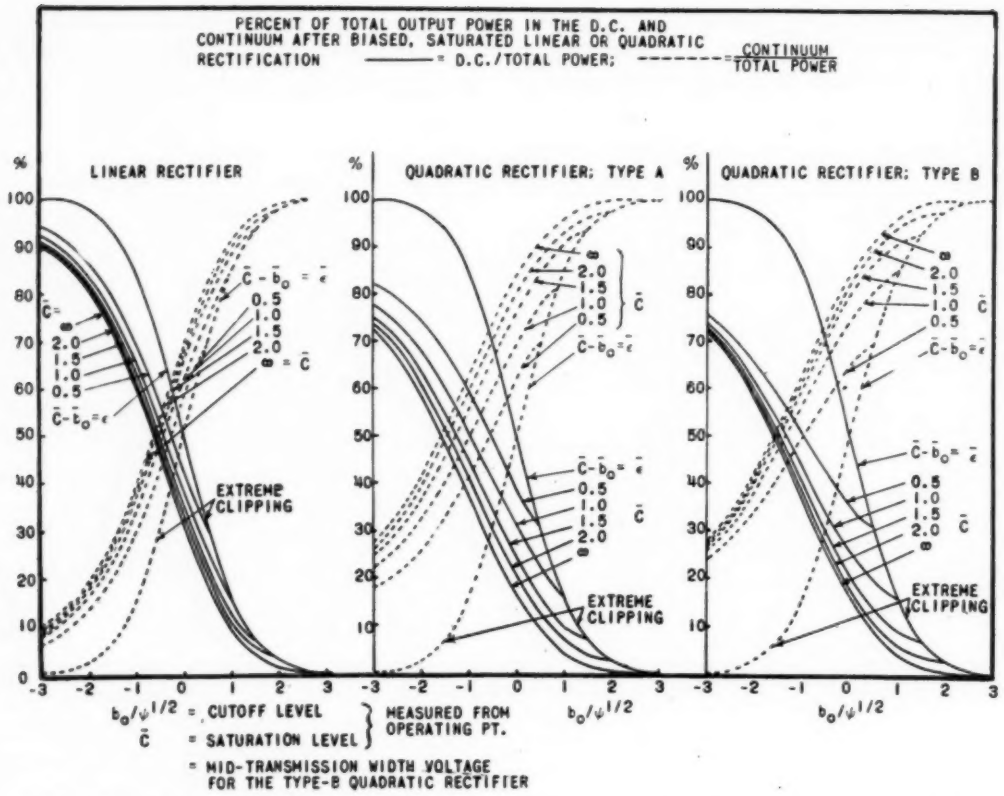
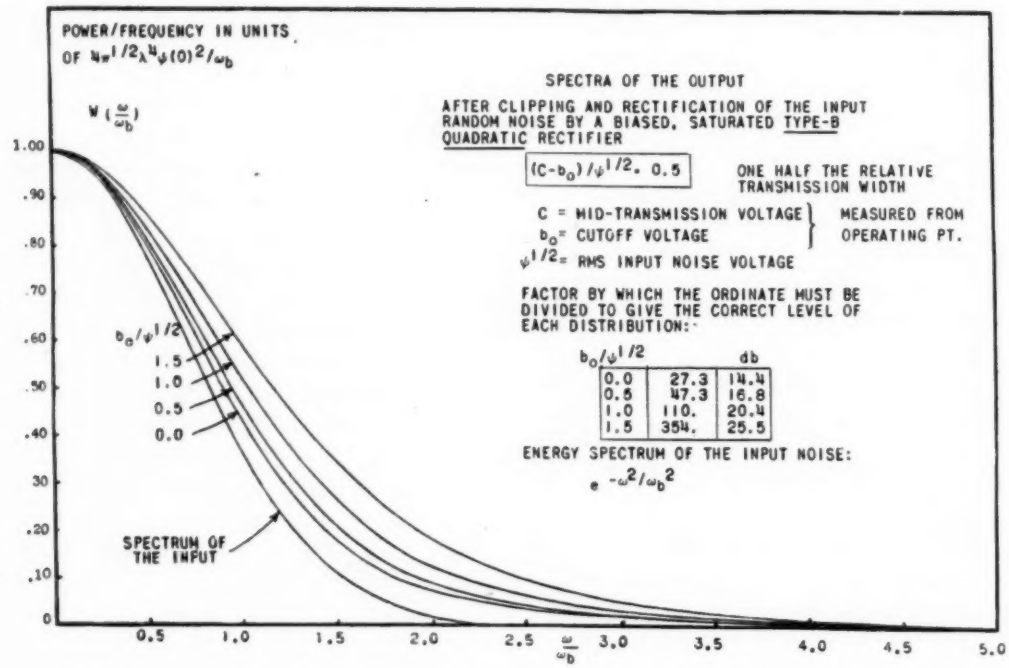

PERCENT OF TOTAL OUTPUT POWER IN THE D.C. AND CONTINUUM AFTER BIASED, SATURATED LINEAR OR QUADRATIC RECTIFICATION
——— = D.C./TOTAL POWER; - - - - - = CONTINUUM / TOTAL POWER
LINEAR RECTIFIER
QUADRATIC RECTIFIER; TYPE A
QUADRATIC RECTIFIER; TYPE B
EXTREME CLIPPING
$b_0/\psi^{1/2}$ = CUTOFF LEVEL
$\bar{C}$ = SATURATION LEVEL
MEASURED FROM OPERATING PT.
= MID-TRANSMISSION WIDTH VOLTAGE FOR THE TYPE-B QUADRATIC RECTIFIER

FIG. 23. Fraction of the total power in the d.c. and continuous portions of the output for the linear, Type-A, and Type-B quadratic detectors.


POWER/FREQUENCY IN UNITS OF $4\pi^{1/2}\lambda^4\psi(0)^2/\omega_b$
$W(\frac{\omega}{\omega_b})$
SPECTRA OF THE OUTPUT
AFTER CLIPPING AND RECTIFICATION OF THE INPUT RANDOM NOISE BY A BIASED, SATURATED TYPE-B QUADRATIC RECTIFIER
$(C-b_0)/\psi^{1/2}$. 0.5
ONE HALF THE RELATIVE TRANSMISSION WIDTH
C = MID-TRANSMISSION VOLTAGE
b_0 = CUTOFF VOLTAGE
MEASURED FROM OPERATING PT.
$\psi^{1/2}$ = RMS INPUT NOISE VOLTAGE
FACTOR BY WHICH THE ORDINATE MUST BE DIVIDED TO GIVE THE CORRECT LEVEL OF EACH DISTRIBUTION:
ENERGY SPECTRUM OF THE INPUT NOISE:
$e^{-\omega^2/\omega_b^2}$
SPECTRUM OF THE INPUT

$b_0/\psi^{1/2}$		db
0.0	27.3	14.4
0.5	47.3	16.8
1.0	110.	20.4
1.5	354.	25.5

FIG. 24. Output spectra for broad band input noise (I) from a biased and saturated Type B quadratic detector.

reduce the spread in the spectrum of clipped noise and when the band width of the filter is narrow compared with the spectrum, randomness is restored and the effects of clipping are no longer apparent.

From Fig. 12 for the low frequency output of narrow band noise ($l=0$) it appears that when clipping occurs near the center of the wave, i.e. $\bar{b}_0=0$, the spread is greatest, although the level is well down. In the case of the first harmonic output zone, Fig. 15, $l=1$, corresponding to a band centered about f_0, the reverse seems to be true, and the greatest spread takes place for the larger values of cut-off $\bar{b}_0$, although this spread is not so pronounced. The reason for this may be appreciated by reference to Eq. (3-23) which shows that when $l=1$, the leading term is generated by taking $n=1$; on the other hand, for the low frequency continuum $l=0$, and the leading term starts with $n=2$; in the former instance we have the spectral contribution $\exp(-\omega^2/\omega_b^2)$, in the latter, $\exp(-\omega^2/2\omega_b^2)$, which suggests why the latter in many cases yields the greater spread. We may remark also that in the neighborhood of spectral "tails," well away from the center of the distribution, the spread is always greatest for extreme clipping.

Figure 11 illustrates the spectral spread after half-wave rectification with saturation, and shows the effects either of decreasing the input-noise voltage or of using tubes whose dynamic paths cover a progressively wider range of voltages. In either case, as the relative transmission width is increased, greater spectral spread obtains and at a higher level, until the limit of saturationless clipping is reached. The level rises because more of the incoming wave is transmitted, and again it is seen that clipping alone due to the tube cut-off results in the greatest spectral spread and at the highest (relative) output level.*

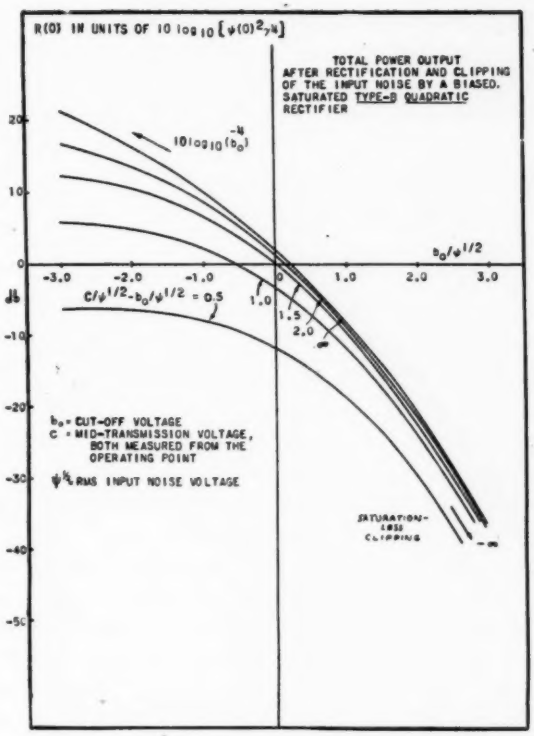


FIG. 25. Total power output of a Type B quadratic rectifier for input noise of classes I–III.

Examples of symmetrical clipping are shown in Fig. 10, in which the rectifier is biased so that the operating point lies half-way between cut-off and saturation, i.e., $b_0=C$. For broad band and semi-broad band noise little spreading of the spectrum results, even when the relative transmission width is greatly decreased, but when the incoming disturbance is narrow band (III), the low frequency ($l=0$) output vanishes. In fact, all even harmonics $l=0, 2, 4, \cdots$ are missing when the clipping is symmetrical. This effect may also be observed in Fig. 21 where the output power goes to zero. As we have mentioned before, amplification in practice nearly always results in approximately symmetrical clipping at both top and bottom of the wave; these results show that the effect is small for input spectra of types I and II, but may be serious in certain output zones of class III waves.

Figures 7, 9, 14, and 17 illustrate instances for which the relative transmission width is in-

* At first glance this may appear odd, since we would expect the additional clipping from overload to spread the spectrum even more. A comparison of the correlation functions for extreme clipping when $b_0=0$, Eq. (6-10), and for half-wave rectification without saturation ($b_0=0$, $C=\infty$), Eq. (4-12), shows that $R(t)_{\text{ext}}\sim r+r^3/6+3r^5/40+\cdots$, while $R(t)_{\text{no sat}}\sim r+r^2/\pi+r^4/12\pi+\cdots$. It is the greater magnitude of the second term, r^2/π *vs.* $r^3/6$, that accounts for the greater spread in the instance of no saturation. It should be observed, however, that a comparison of spectra in the neighborhood of the "tail," *well away from the center of the band*, indicates that the reverse is true. A similar comparison may be made for the quadratic detector with the help of (5-14).

finitesimally small, producing extreme or "superclipping." We see that superclipping spreads the spectrum most when the dynamic path is centered well away from the operating point, i.e., $|b_0| \gg 0$, where the curvature is greatest. (An exception to this is the case $l=0$, Fig. 14.) The spectral level, of course, is extremely low.

(b) Power

As we have demonstrated in Section 3, the power output of our non-linear device, unlike the spectrum of the output, does not depend on the spectrum of the input disturbance, but only on the root-mean-square input-noise voltage. Figure 19 shows the d.c. output power as a function of cut-off voltage, measured from the operating point. As b_0 becomes more positive less of the wave is transmitted and hence the direct current drops and approaches the limiting value for saturationless clipping, which depends only on b_0 and not on the transmission width $(C-b_0)$. On the other hand, when b_0 is negative, the positive part of the incoming disturbance is progressively more and more limited because of saturation. Then the d.c. output rises and approaches a fixed value asymptotically, since it cannot exceed some finite value which depends on the transmission width of the rectifier. For no overloading $(\bar{C}\to\infty)$, this finite limit is removed and represents chiefly the voltage supplied by the tube bias, which is needed to shift the operating point further into the transmission region of the dynamic path.

The powers in the continuum appear in Figs. 20, 21, and 22. Here it is interesting to observe that for symmetrical clipping in the cases of class I and II input spectra, the continuum level is highest, falling off for values of $|b_0|$ in excess of $b_0=-C$. This may be explained by the fact that when the transmission width is centered about the incoming wave more of the wave is effectively passed than if only the tops or bottoms are transmitted. When disturbances having class III spectra are considered, this is seen to be true for the first and higher odd harmonic regions $l=1, 3, 5, \cdots$, *vide* Fig. 22, but for values of $b_0=-C$, there is no low frequency or for that matter, no even-harmonic output, because, with the exception of a d.c. component, only odd-order modulation products are generated. Curves of total output power are given in Fig. 18. Figure 23 shows the percentage distribution of the total output power that appears in the continuum and in the d.c. component. The quadratic responses, it may be noted, yield a somewhat greater fraction of continuum power than does the linear detector, and of these the Type B characteristic yields the most, for a given cut-off and transmission width.

We expressly wish to thank Professor J. H. Van Vleck for his encouragement in the preparation of this manuscript, Mr. S. O. Rice, with whom we have had many friendly and stimulating discussions on these topics, and Miss Eleanor Pressly, who performed most of the computations.

APPENDIX

In Section III we gave the results of the evaluation of (4-2) for the biased and saturated linear detector. A similar procedure is used to evaluate (5-4) for the quadratic rectifier.

Let us consider only the first term in the integrand of (4-2), since the second may be treated in a similar fashion. When $n=0$, the imaginary part contributes only through the pole at the origin, i.e.,

$$(h_{0,0})_{b_0}=\frac{-\beta}{2\pi}\left\{\int_{\mathbf{C}}\cos b_0 z\exp(-\psi z^2/2)dz/z^2+\pi b_0\right\}, \quad \text{(A-1)}$$

since the effective contour is a semicircle. The real part, however, is somewhat more difficult to evaluate. By adding $-1/z^2$ to the integrand we may remove the singularity and hence replace $\mathbf{C}$ by a path entirely along the real axis. It is easily shown that $\int_{\mathbf{C}}z^{-2}dz$ vanishes. We now have an integral of the form

$$I_1=\int_{-\infty}^{\infty}[\cos b_0 z\exp(-\psi z^2/2)-1]dz/z^2. \quad \text{(A-2)}$$

Differentiating (A-2) with respect to b_0 we obtain

$$\frac{\partial I_1}{\partial b_0}=-2\int_0^{\infty}\sin b_0 z\exp(-\psi z^2/2)dz/z$$
$$=-\pi\Theta(b_0/(2\psi)^{\frac{1}{2}}), \quad \text{(A-3)}$$

where Θ is defined in (4-8). Integration of (A-3) by parts yields

$$I_1(b_0)-I_1(0)$$
$$=-\pi\{b_0\Theta b_0+(2\psi/\pi)^{\frac{1}{2}}(\exp(-b_0^2/2\psi)-1)\}. \quad \text{(A-4)}$$

Now from (A-2) we observe that

$$I_1(0)=\int_{-\infty}^{\infty}[\exp(-\psi z^2/2)-1]dz/z^2; \quad \text{(A-5)}$$

hence we may write

$$\frac{\partial I_1(0)}{\partial\psi}=-\int_0^{\infty}\exp(-\psi z^2/2)dz=-(\pi/2\psi)^{\frac{1}{2}}. \quad \text{(A-6)}$$

Finally we have

$$I_1(0)=-(2\pi\psi)^{\frac{1}{2}}. \quad \text{(A-7)}$$

The constant of integration is zero here, since for $\psi=0$ $I_1(0)=0$. From (A-4) and (A-7) it follows that

$$(h_{0,0})_{b_0}=\beta\{b_0(\Theta b_0-1)/2+\psi^{\frac{1}{2}}\phi b_0\}, \quad \text{(A-8)}$$

with a similar result when b_0 is replaced by C. Equation (4-3) may be obtained from (4-2) and the above. For $n=1$, we have

$$(h_{0,1})b_0 = -\frac{\beta}{2\pi}\left[\pi i - i\int_{-\infty}^{\infty} \sin b_0 z \exp(-\psi z^2/2)dz/z\right] = \beta i(\Theta b_0 - 1)/2. \quad \text{(A-9)}$$

from which (4-4) is established in a similar fashion. When $h \geqq 2$, the contour C is replaced by a path along the real axis. Then to determine $h_{0,n}$ we need the following results, which are readily established by differentiation with respect to λ:

$$\int_0^\infty z^{2p}\cos\lambda z \exp(-\psi z^2/2)dz = (\pi/2)^{\frac{1}{2}}(-1)^p\psi^{-p-\frac{1}{2}}H_{2p}(\lambda/\psi^{\frac{1}{2}})\exp(-\lambda^2/2\psi) = \pi(-1)^p\psi^{-p-\frac{1}{2}}\phi^{(2p)}(\lambda/\psi^{\frac{1}{2}}), \quad \text{(A-10)}$$

and

$$\int_0^\infty z^{2p-1}\sin\lambda z \exp(-\psi z^2/2)dz = (\pi/2)^{\frac{1}{2}}(-1)^p\psi^{-p}H_{2p-1}(\lambda/\psi^{\frac{1}{2}})\exp(-\lambda^2/2\psi), \quad \text{(A-11)}$$
$$= \pi(-1)^p\psi^{-p}\phi^{(2p-1)}(\lambda/\psi^{\frac{1}{2}}), \quad p=0, 1, 2\cdots$$

where $H_j(x_0)$ is an Hermitian polynomial of the jth order,[13] cf. Eqs. (4–9). Here we note also that[14]

$$H_{-1}(x_0) \equiv (2\pi)^{\frac{1}{2}}\phi^{(-1)}(x_0) = (\pi/2)^{\frac{1}{2}}\exp(+x_0^2/2)\Theta(x_0/2^{\frac{1}{2}}), \quad \text{(A-12)}$$

and that

$$H_{2p}(0) = (-1)^p(2p)!/2^p p!, \qquad H_{2p-1}(0) = 0, \quad p=0, 1, 2, \cdots. \quad \text{(A-13)}$$

The explicit form for $\phi^{(j)}(x_0)$ is useful, namely

$$\phi^{(j)}(x_0) = \exp(-x_0^2/2)\sum_{p=0}^{j/2,\,(j-1)/2}(-1)^{p+1}j!x_0^{j-2p}/p!(j-2p)!2^p(2\pi)^{\frac{1}{2}}. \quad \text{(A-14)}$$

The functions $\phi^{(j)}$ satisfy the recurrence relation

$$\phi^{(j+1)}(x_0) = -[x_0\phi^{(j)}(x_0) + j\phi^{(j-1)}(x_0)], \quad j=1, 2, \cdots. \quad \text{(A-15)}$$

[13] See, for example, T. C. Fry, *Probability and Its Engineering Uses* (D. Van Nostrand Co., Inc., 1928), pp. 251–254, and for tables of $\phi^{(0)}-\phi^{(6)}$ consult pages 456, 457.

[14] B. O. Pierce, *A Short Table of Integrals* (Ginn and Company, New York, 1929), contains a table of Θ.

Calendar of Meetings

October

25 American Mathematical Society, New York, New York

November

1– 2 Society of Rheology, New York, New York
2 American Physical Society (New York State Section), Saratoga Springs, New York
7– 9 Conference on Electrical Insulation, Division of Engineering and Industrial Research, National Research Council, Baltimore, Maryland
14–16 Acoustical Society of America, Chicago, Illinois
29–30 American Physical Society, Minneapolis, Minnesota

December

2– 6 American Society of Mechanical Engineers, New York, New York
5– 7 Electron Microscope Society of America and the American Society for X-Ray and Electron Diffraction, Pittsburgh, Pennsylvania (Joint Meeting)
17 Institute of Aeronautical Sciences, New York, New York
26–28 Geological Society of America, Chicago, Illinois

January

23–25 American Physical Society, New York, New York

The Flow of Gases in Pipes at Low Pressures

GORDON P. BROWN AND ALBERT DINARDO
National Research Corporation, Boston, Massachusetts
AND
GEORGE K. CHENG AND THOMAS K. SHERWOOD
Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 9, 1946)

Data are presented on the flow of air at low pressures in copper pipes of radius 0.795 and 1.30 cm; iron pipes of radius 2.64, 5.12, and 10.1 cm; and for hydrogen in a copper pipe of radius 1.30 cm. Data are also given for the resistance to flow of air in 1½- and 3-inch nominal pipe size elbows, for 1½-inch Kinney bellows valves, and for short sections of 1-inch nominal pipe size iron pipe connecting large gas chambers. The data on straight iron and copper pipes, together with the data from the literature on glass capillaries, are correlated by the introduction of a correction factor F in Poiseuille's equation. F is correlated graphically in terms of a dimensionless group representing the ratio of mean free path to pipe radius, with separate curves for iron pipe and for copper and glass pipes or capillaries.

INTRODUCTION

INDUSTRIAL processes operating at high pressure have been common for a generation, but it is only during the last several years that the other extreme of operating pressures has come to be used in a variety of large scale operations. Pressures as low as a few microns are employed in the manufacture of electronic equipment, production of magnesium metal from dolomite, the drying of blood plasma and of penicillin, the concentration of vitamins, and the coating of lenses used in military optical equipment.[1] Moderately low pressures are being used in concentration of fruit juices and other food products.

The necessary vacuum is obtained by the use of mechanical pumps, multi-stage steam jets, or large oil diffusion pumps. No matter how good the vacuum at the pump, it is evident that the pressure in the processing equipment may be unnecessarily high and perhaps excessive if the piping and connections are not designed to handle the gas or vapor flow with an economically small pressure loss due to friction. The proper design of conduits for gas flow at low pressures introduces new problems. At ordinary pressures it is customary to employ the combination of the Fanning equation and a friction factor graph if the flow is turbulent (Reynolds number greater than about 2100) and to use Poiseuille's law if the flow is viscous or laminar (Reynolds number below about 1000). At low pressures (below 1.0 mm Hg) the Reynolds number is probably always so low as to indicate laminar flow, but Poiseuille's law does not apply since the effective viscosity, even in industrial pressure ranges, may be less than one percent of the gas viscosities as tabulated in the standard handbooks.

[1] Anonymous, Fortune **20**, No. 2, 147 (August, 1944).

FLOW EQUATIONS—STREAMLINE FLOW

The differential and integrated forms of Poiseuille's law for round tubes may be written

$$Q=-\frac{\pi g_c R^4 p}{8\mu}\frac{dp}{dL}=\frac{\pi g_c R^4}{16\mu L}(p_1^2-p_2^2). \qquad (1)$$

The derivation of this equation, frequently quoted in textbooks, is based on the assumption that the coefficient of viscosity, μ, is a constant and that the velocity of gas in contact with the tube wall is zero. Gas viscosities, as predicted by the kinetic theory, are essentially constant over a wide range of moderate and high pressures. Consequently, the failure of the equation at low pressures is explained as caused by the failure of the assumption of no slip at the wall rather than on the basis of a variable viscosity.

SLIP FLOW

The derivation of Poiseuille's equation has been repeated by Kundt and Warburg[2,3] with

[2] A. Kundt and E. Warburg, Pogg. Ann. **155**, 337, 525 (1875).
[3] E. Warburg, Wied. Ann. **159**, 399 (1876).

TABLE I. Systems of units.

	Consistent c.g.s. technical	Consistent English	Micron c.g.s.	Micron English
Q	$\frac{cm^3}{sec.}\times\frac{g}{cm^2}$	$\frac{ft.^3}{sec.}\times\frac{lb.}{ft.^2}$	$\frac{micron}{sec.}\times cm^3$	$\frac{micron\times ft.^3}{sec.}$
μ*	$\frac{g}{sec.\times cm}$ (poises)	$\frac{lb.}{sec.\times ft.}$	poises	$\frac{lb.}{sec.\times ft.}$
R_G	84800	1543	6.24×10^{-7}	5.56×10^{5}
R, L, λ	cm	ft.	cm	ft.
p_1, p_2, p_m	g/cm²	lb./ft.²	microns	microns
ρ_1	$\frac{g}{cm^3}\times\frac{cm^2}{g}$	$\frac{lb.}{ft.^3}\times\frac{ft.^2}{lb.}$	$\frac{g}{cm^3\times microns}$	$\frac{lb.}{ft.^3\times microns}$
f, π		No units		
g_c	981	32.2	1.33	0.0896

* Viscosity in lb./(sec.)(ft.) = 0.0672 × viscosity in poises = 0.000672 × viscosity in centipoises.

the modification that the velocity of the gas in contact with the wall is taken as u_0 instead of zero when the mean free path is comparable with the tube radius. The resulting tangential force F' on the wall is assumed to be proportional to u_0 and to the surface area

$$F' = ku_0A_w, \tag{2}$$

where the proportionality constant k is termed the coefficient of external friction. Balancing the forces on a differential length dL of tube,

$$-\pi R^2 g_c dp = F' = 2\pi R k u_0 dL, \tag{3}$$

whence

$$u_0 = \frac{Rg_c}{2k}\frac{dp}{dL}. \tag{4}$$

This last is used as one limit of u in the usual integration to obtain:

$$Q = \frac{\pi g_c R^4}{16\mu L}(p_1^2 - p_2^2)\left(1+\frac{4\mu}{kR}\right). \tag{5}$$

Maxwell[4] suggested that a fraction f of the gas molecules striking the walls of a vessel would be diffusely reflected, and that the fraction $1-f$ would be specularly reflected. With this concept, it can be shown by the kinetic theory that,

$$\frac{\mu}{k} = \left(\frac{2}{f}-1\right)\lambda_m, \tag{6}$$

which may be introduced into (5) to give,

$$Q = \frac{\pi g_c R^4}{16\mu L}(p_1^2 - p_2^2)\left[1+4\left(\frac{2}{f}-1\right)\frac{\lambda_m}{R}\right], \tag{7}$$

where λ_m is the mean free path of the gas molecules at the mean pressure. Since the coefficient of external friction, k, has been found[2] to be proportional to gas pressure, its value and that of λ should be evaluated at the mean gas pressure. From the kinetic theory and Maxwell's distribution of molecular velocities, the mean free path can be shown to be given by

$$\lambda_m = \left(\frac{\pi}{2\rho_1 g_c}\right)^{\frac{1}{2}}\cdot\frac{\mu}{p_m},$$

where ρ_1 is the gas density at unit pressure, M/R_GT. Thus, (7) becomes

$$Q = \frac{\pi g_c R^4}{16\mu L}(p_1^2 - p_2^2)\left[1+4\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\left(\frac{2}{f}-1\right)X\right], \tag{8}$$

where X represents the group $\mu/p_mR(\rho_1 g_c)^{\frac{1}{2}}$. Inspection of (7) shows that this equation for "slip flow" is essentially Poiseuille's equation (1) multiplied by a correction factor which becomes more important as the ratio of mean free path to radius increases and as f becomes small (smooth walls). Equation (8) is a restatement of (7) in terms of more directly useful variables.

MOLECULAR FLOW

At very low pressures the internal friction in the moving gas becomes small because the mean free path of the gas molecules becomes large in relation to the dimensions of the conduit. Under these conditions the collisions between molecules are rare as compared with the frequency of collisions with the walls. By employing the kinetic theory in a balance of momentum transfer and pressure difference in a pipe section, there is obtained[5]

$$Q = \left(\frac{\pi}{2}\right)^{\frac{1}{2}}\left(\frac{2}{f}-1\right)\frac{R^3 g_c}{L}\frac{(p_1-p_2)}{(\rho_1 g_c)^{\frac{1}{2}}}. \tag{9}[6]$$

[4] J. C. Maxwell, *The Scientific Papers of James Clerk Maxwell* (University Press, Cambridge, 1890), Volume 2, p. 708.

[5] R. A. Millikan, Phys. Rev. **21**, 217 (1923).

[6] M. Knudsen [Ann. d. Physik **28**, 75, 999 (1909)] taking f as unity, obtained the same form with a constant 70 percent greater.

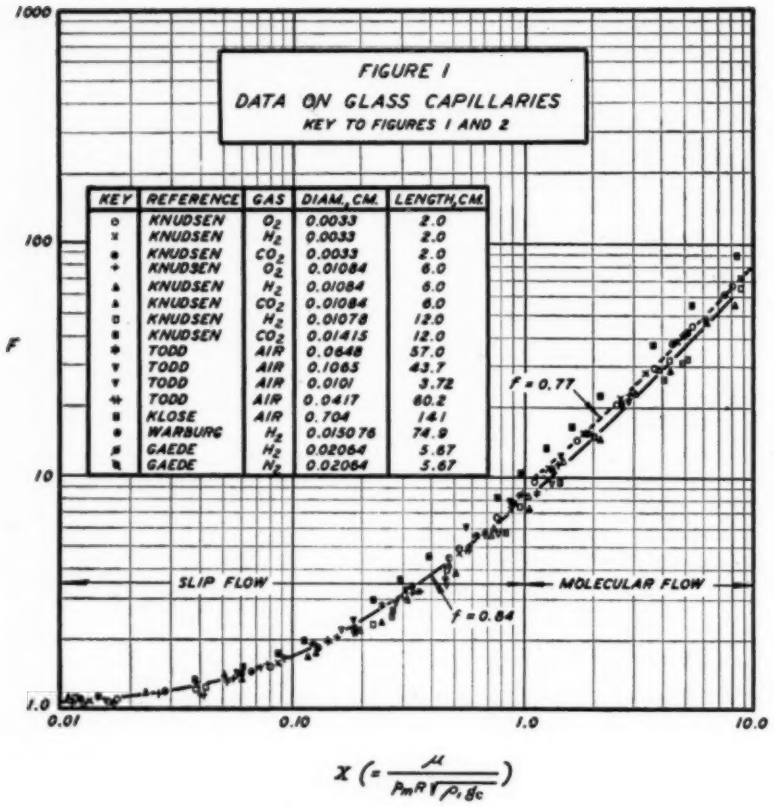


FIGURE 1
DATA ON GLASS CAPILLARIES
KEY TO FIGURES 1 AND 2
KEY | REFERENCE | GAS | DIAM., CM. | LENGTH, CM.
KNUDSEN | O_2 | 0.0033 | 2.0
KNUDSEN | H_2 | 0.0033 | 2.0
KNUDSEN | CO_2 | 0.0033 | 2.0
KNUDSEN | O_2 | 0.01084 | 6.0
KNUDSEN | H_2 | 0.01084 | 6.0
KNUDSEN | CO_2 | 0.01084 | 6.0
KNUDSEN | H_2 | 0.01078 | 12.0
KNUDSEN | CO_2 | 0.01415 | 12.0
TODD | AIR | 0.0648 | 57.0
TODD | AIR | 0.1065 | 43.7
TODD | AIR | 0.0101 | 3.72
TODD | AIR | 0.0417 | 60.2
KLOSE | AIR | 0.704 | 141
WARBURG | H_2 | 0.013076 | 74.9
GAEDE | H_2 | 0.02064 | 5.67
GAEDE | N_2 | 0.02064 | 5.67
1000
100
10
1.0
F
f = 0.77
f = 0.84
SLIP FLOW
MOLECULAR FLOW
0.01
0.10
1.0
10.0
$x\ (=\frac{\mu}{P_m R\sqrt{\rho_1 g_c}})$

The nature of the motion of the gas at very low pressures is termed "molecular flow" and Eq. (9) is a quantitative expression for the rate of molecular flow.

Inspection of the "slip flow" Eq. (8) reveals that for very low pressures the second of the two additive terms becomes large in relation to unity, and the equation reduces to (9). Equation (8) is more than an equation for the slip flow region; at high pressures it reduces to Poiseuille's equation (1), and at very low pressures it reduces to the molecular flow Eq. (9).

ENTRANCE AND EXIT LOSSES

For flow of gases through an orifice at very low pressures ("molecular effusion") Knudsen[6] has derived the following equation from the kinetic theory

$$Q=\left(\frac{\pi}{2\rho_1 g_c}\right)^{\frac{1}{2}} R^2(p_1-p_2)g_c. \qquad (10)$$

Comparison of this with Eq. (9) for molecular flow indicates that molecular flow in both orifices and tubes can be expressed by the relation,

$$Q=K\left(\frac{\pi}{2\rho_1 g_c}\right)^{\frac{1}{2}} R^2(p_1-p_2)g_c, \qquad (11)$$

where $K=1$ for $L/R=0$ and approaches

$$K=\frac{\pi}{2}\left(\frac{2}{f}-1\right)\frac{R}{L}$$

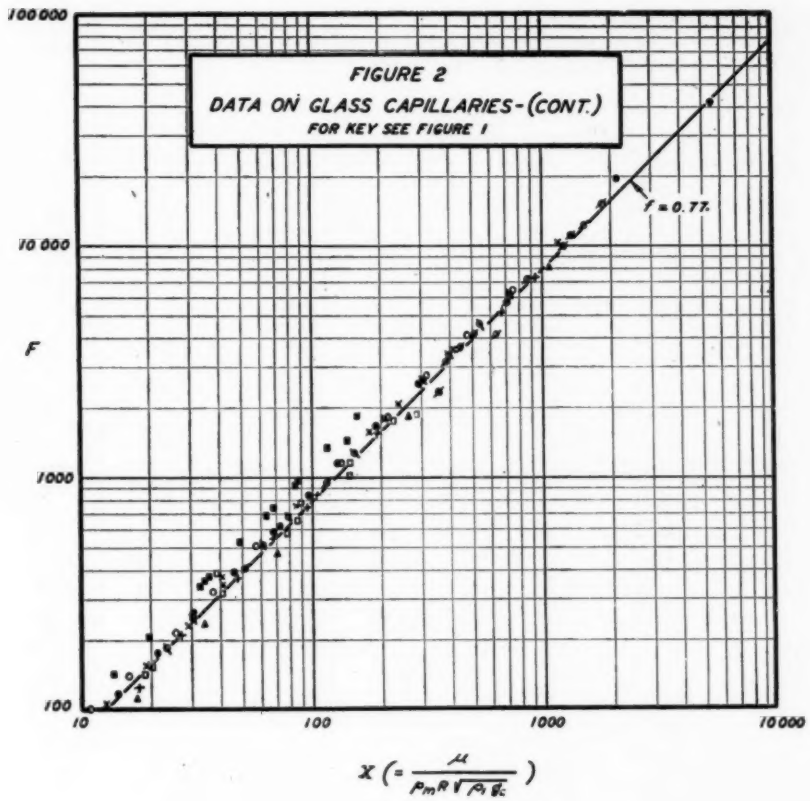
as L/R approaches infinity. Clausing[7] gives values of K as a function of R/L, varying from 1.0 to

$$\frac{8}{3}\frac{R}{L}.$$

The latter limit corresponds to a constant value of f of 0.74.

In streamline flow the entrance loss may be taken as equal to the velocity head, u^2/g_c. A graph of pressure drop *vs.* Q for an orifice or for the entrance to a tube may be represented in part by two straight lines on logarithmic co-

[7] P. Clausing, Ann. d. Physik [5*F*] **12**, 961 (1932).



ordinates. At low pressures the relation is indicated by (11) and the slope is unity. In streamline flow

$$Q=\pi R^2 p_2\left(\frac{R_G T(p_1-p_2)g_c}{p_m M}\right)^{\frac{1}{2}} \qquad (12)$$

and there is a family of lines having slopes of 2. The data cross from the first line to one of the second type as the pressure increases, and there is some indication[8] that the experimental points pass through an inflection and a minimum in so doing.

UNITS

The various equations given in the preceding sections are valid in any consistent set of technical units; if g_c is taken as unity, they are valid in absolute units. The common mixture of pressure in microns Hg with English units is permissible if an appropriate value of g_c is employed. The proper use of the four technical systems is indicated in Table I.

SUMMARY OF THEORY

Although the flow of gases at low pressures may be described as streamline flow, slip flow, or molecular flow, the single Eq. (8) is seen to represent the theory for all three regions. Alternatively, Poiseuille's equation (1) may be employed, with the right-hand side multiplied by a correction factor F:

$$S=\frac{Q}{(p_1-p_2)}=\frac{F\pi R^4 p_m g_c}{8\mu L}, \qquad (13)$$

where

$$F=1+4\left(\frac{\pi}{2}\right)^{\frac{1}{2}}(2/f-1)X. \qquad (14)$$

Here μ is the viscosity of the gas at atmospheric pressure, and the factor F may be looked upon as the ratio of μ to the effective viscosity applicable to Poiseuille's equation at low pressures.

[8] G. K. Cheng, Sc.D. Thesis in Chemical Engineering, Massachusetts Institute of Technology, 1945.

TABLE II. Experimental data on flow of gases in copper and iron pipes. (p_1 and p_2 represent upstream and downstream pressures in microns; U is the speed in micron cu. ft. per sec. per micron.)

Run	p_1	p_2	U	F	$\frac{\mu}{p_m R(\rho_1 g_c)^{\frac{1}{2}}}$
A. Copper pipe, $R=1.30$ cm, $L=332$ cm, air, using 4-in. oil diffusion pump and Megavac.					
1	4	1	0.020	9.48	1.26
2	5	1	0.0186	7.28	1.05
3	6	2	0.0197	5.78	0.79
4	7	2	0.0195	5.09	0.70
5	8	3	0.025	5.25	0.56
6	10	4	0.022	3.75	0.45
7	11	5	0.026	3.74	0.39
8	12	4	0.026	3.78	0.39
9	14	5	0.027	3.31	0.33
10	18	8	0.029	2.62	0.24
11	20	16	0.028	1.85	0.185
12	28	23	0.038	1.74	0.124
13	30	25	0.043	1.85	0.115
14	38	32	0.049	1.66	0.090
15	53	46	0.066	1.56	0.064
16	54	47	0.066	1.45	0.063
17	64	55	0.073	1.44	0.053
18	70	61	0.075	1.35	0.048
19	77	68	0.082	1.33	0.044
20	79	69	0.088	1.39	0.043
21	90	78	0.087	1.21	0.038
22	90	79	0.090	1.25	0.037
23	120	106	0.114	1.19	0.028
24	143	128	0.132	1.14	0.023
25	159	142	0.143	1.12	0.021
26	162	146	0.151	1.15	0.021
27	167	150	0.152	1.13	0.020
28	188	170	0.166	1.09	0.0186
29	223	203	0.190	1.05	0.0148
30	261	239	0.224	1.05	0.0126
31	283	260	0.247	1.04	0.0116
32	333	311	0.319	1.03	0.0098
33	393	368	0.340	1.05	0.0085
34	588	561	0.502	1.01	0.0055
B. Copper pipe, $R=1.30$ cm, $L=332$ cm, hydrogen, using 4-in. oil diffusion pump and Megavac.					
1	4	3	0.093	15.3	1.66
2	8	6	0.079	6.45	0.83
3	9	7	0.091	6.52	0.73
4	11	8	0.089	5.39	0.61
5	16	12	0.091	3.72	0.41
6	21	16	0.091	2.83	0.31
7	17	15	0.093	3.33	0.36
8	17	15	0.092	3.32	0.36
9	17	15	0.097	3.49	0.36
10	18	16	0.107	3.62	0.34
11	25	22	0.103	2.52	0.26
12	27	24	0.114	2.57	0.23
13	29	26	0.129	2.69	0.21
14	40	36	0.143	2.18	0.153
15	48	43	0.143	1.81	0.138
16	55	50	0.172	1.88	0.110
17	61	55	0.162	1.60	0.100
18	71	65	0.198	1.67	0.085
19	82	75	0.205	1.50	0.074
20	125	116	0.271	1.30	0.048
21	133	124	0.290	1.30	0.044
22	295	284	0.645	1.28	0.020
23	327	314	0.620	1.11	0.0181
24	406	393	0.773	1.11	0.0145
25	595	580	1.08	1.06	0.0093
26	650	635	1.08	1.05	0.0090
27	681	665	1.19	1.02	0.0086
28	758	742	1.27	0.97	0.0077
29	977	960	1.77	1.05	0.0060
30	1017	1000	1.80	1.03	0.0058

Run	p_1	p_2	U	F	$\frac{\mu}{p_m R(\rho_1 g_c)^{\frac{1}{2}}}$
C. Copper pipe, $R=0.795$ cm, $L=127$ cm; air, using N.R.C. oil diffusion pump H-6 and 50 cfm Kinney mechanical pump.					
1	470	130	0.099	1.07	0.0172
2	370	100	0.083	1.14	0.022
3	270	70	0.061	1.15	0.030
4	198	48	0.045	1.18	0.042
5	148	34	0.037	1.31	0.056
6	84	15	0.023	1.48	0.103
7	48	7	0.0168	1.93	0.184
8	29	4	0.0140	2.73	0.31
9	14	2	0.0126	5.06	0.64
10	7	1	0.0120	9.7	1.29
D. Iron pipe, $R=2.64$ cm, $L=274$ cm; air, using N.R.C. oil diffusion pump H-6, and 50 cfm Kinney mechanical pump.					
1	1.35	0.25	0.136	9.45	1.94
2	4.00	0.80	0.159	3.78	0.65
3	7.80	1.60	0.180	2.18	0.33
4	23.20	6.50	0.34	1.31	0.105
5	14.60	3.80	0.26	1.61	0.168
6	41.20	13.0	0.54	1.14	0.057
7	59.9	22.4	0.85	1.17	0.038
8	102.0	57.0	1.36	0.93	0.0195
9	110.0	66.5	1.37	0.89	0.0175
10	136.0	100.0	2.12	1.02	0.0132
11	187	150	2.72	0.92	0.0092
12	248	211	3.79	0.94	0.0068
13	320	290	4.81	0.90	0.0051
14	385	352	5.94	0.89	0.0041
15	447	412	6.69	0.88	0.0036
16	558	528	9.91	1.04	0.0029
17	620	588	10.6	1.00	0.0026
18	769	731	11.7	0.89	0.0021
19	880	842	13.0	0.86	0.0018
20	0.64	0.13	0.164	25.4	4.20
21	1.56	0.30	0.158	9.65	1.67
22	2.61	0.50	0.160	5.9	1.00
23	3.40	0.68	0.167	47	0.76
24	0.45	1.10	0.172	36.2	5.65
25	1.00	0.22	0.179	16.7	2.54
26	0.60	0.10	0.144	23.4	4.44
27	2.05	0.40	0.170	7.96	1.27
E. Iron pipe, $R=5.12$ cm, $L=457$ cm; air, using N.R.C. oil diffusion pump E-10 and 100 cfm Kinney mechanical pump.					
1	101	37	10.5	1.02	0.0116
2	73	24	8.10	1.11	0.0163
3	60.9	17.4	6.26	1.07	0.0202
4	45.0	11.5	4.57	1.08	0.0283
5	35.0	8.2	3.65	1.13	0.037
6	23.8	4.6	2.58	1.21	0.056
7	17.6	3.1	2.02	1.30	0.077
8	13.0	2.2	1.71	1.51	0.105
9	8.0	1.1	1.19	1.74	0.174
10	5.8	0.8	1.04	2.11	0.242
11	3.99	0.44	0.77	2.36	0.366
12	2.5	0.30	0.70	3.35	0.57
13	1.14	0.11	0.60	6.4	1.27
14	0.52	0.06	0.62	14.2	2.76
15	0.29	0.02	0.57	24	5.0
16	1.80	0.20	0.65	6.5	0.80
17	435	413	50	0.79	0.0019
18	319	289	34	0.75	0.0026
19	197	164	27	0.99	0.0044
20	160	118	17.6	0.85	0.0058
21	95.0	34.0	9.3	0.97	0.0124

TABLE II (*Continued*).

Run	p_1	p_2	U	F	$\frac{\mu}{p_m R(\rho_1 g_c)^{\frac{1}{2}}}$
F. Iron pipe, $R=10.1$ cm, $L=1040$ cm; air, using N.R.C. oil diffusion pump E-10 and 100 cfm Kinney mechanical pump.					
1	84	72	62	0.79	0.0052
2	80	67	54	0.73	0.0055
3	48	30	34	0.88	0.0104
4	32.5	12	19.9	0.91	0.0184
5	21.9	5.7	13.2	0.96	0.029
6	16.9	4.2	11.2	0.97	0.038
7	11.5	2.7	8.1	1.14	0.057
8	5.6	1.0	4.6	1.40	0.123
9	4.0	0.6	3.6	1.56	0.176
10	2.1	0.3	2.6	2.20	0.34
11	1.35	0.24	2.5	3.1	0.51
12	1.1	0.20	2.3	3.6	0.62
13	0.38	0.12	2.6	10.4	1.62
14	0.25	0.05	2.06	13.8	2.70
15	0.15	0.05	3.5	35	4.05

CORRELATION OF DATA FROM THE LITERATURE

Theoretical equations are often of value in suggesting dimensionless groups to be used as a basis for an empirical correlation, even though the data and theory do not agree quantitatively. In the present instance Eqs. (8) and (14) suggest the probability of a relation between F, X, and f. Figures 1 and 2 show the result of plotting F *vs.* X for the various data in the literature[3, 6, 9–11] on flow of air, H_2, O_2, N_2, and CO_2 at low pressures in round glass tubes of small diameter. The correlation for the five gases is excellent, in spite of the 200-fold variation in tube radius. The solid line shown on Fig. 2 represents

$$F=8.0X$$

corresponding to $f=0.77$. In the range of higher pressures shown in Fig. 1 ($X<1.0$), the solid line best representing the data corresponds to $f=0.84$.

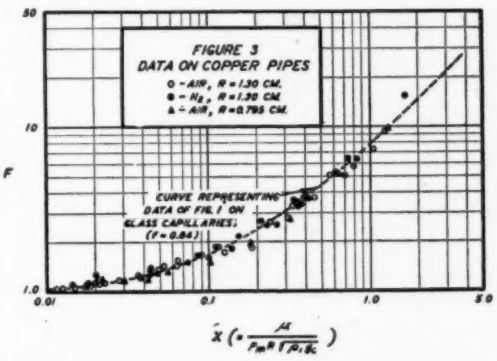
Figure 2 indicates measured values of F as large as 40,000. This means that the flow at low pressures reaches rates as great as 40,000 times that which would be predicted by Poiseuille's equation, i.e., the effective viscosity of the gas at low pressures may be as small as one forty-thousandth the viscosity at one atmosphere.

[9] G. W. Todd, Univ. Durham Phil. Soc. Proc. 6, Part I, 8 (1920).
[10] W. Klose, Physik. Zeits. 31, 503 (1930).
[11] W. Gaede, Ann. d. Physik 41, 289 (1913).

EXPERIMENTAL

Although the correlation of data from the literature is satisfactory, it is evident that extrapolation from glass tubes 7 mm and less in radius to iron pipes several inches in diameter may be considered to be questionable. Accordingly, new data have been obtained for copper and iron pipe $\frac{5}{8}$ to 8 in. nominal diameter. Data were also obtained on $1\frac{1}{2}$- and 3-in. elbows and a $1\frac{1}{2}$-in. Kinney high vacuum bellows valve.

The first experiments were carried out using a copper tube 1.30 cm radius and 332 cm between taps to two McLeod gauges. Carefully dried air was admitted through a vacuum reducing valve to a 5-liter vessel from which it passed through an open mercury seal to the test pipe. After a steady state had been reached the mercury seal was closed and the rate of increase of pressure in the 5-liter vessel measured by means of a McLeod gauge. The gas feed rate was calculated from the rate of pressure increase and the volume of the vessel. Pumping was accomplished by means of a 4-inch oil diffusion pump and a Megavac pump in series. A thermocouple gauge



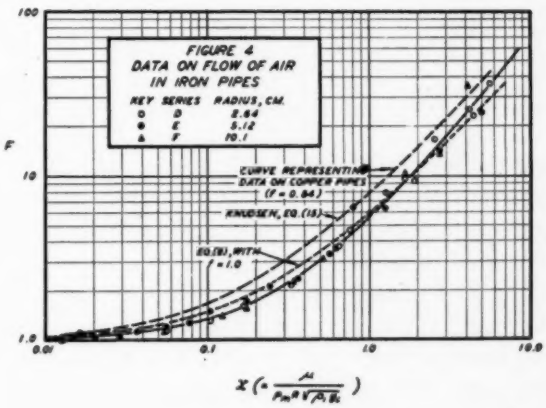


TABLE III. Flow of air in elbows and valves.

Run	p_1	p_2	U	Re	L_e
A. Eight 1½-in. elbows and 277 cm straight 1½-in. nominal pipe size (R =2.04 cm) iron pipe between pressure taps.					
1	4.4	0.6	0.089	0.025	−1.7
2	2.16	0.26	0.074	0.0104	−0.5
3	7.1	1.1	0.100	0.045	−1.9
4	22.1	4.1	0.135	0.179	−0.6
5	49.0	11.0	0.215	0.604	0.6
6	81.0	21.0	0.36	1.58	0
7	120.0	38.0	0.51	3.08	0.5
8	144.5	63.5	0.70	4.2	0
9	150.0	108.0	0.84	2.6	0.1
10	238.0	180.0	1.28	5.5	0.3
11	300.0	230.0	1.50	7.75	0.9
12	410.5	343.5	2.5	12.6	−0.5
13	504.5	429.5	3.3	18.3	−1.0
14	637.5	562.5	4.3	24	−1.1
15	790.0	690.0	4.5	33	0.2
16	980.0	870.0	5.4	44	0.7
17	1016.5	919.5	6.1	44	−0.2
18	706.0	620.0	4.9	31	−1.4
19	1201	1099	7.0	53	0.2
20	1971	1849	9.5	85	1.8
21	3070	2930	12.4	129	4.3
22	3978	3821	14.7	170	2.5
23	3677	3523	15.2	173	4.1
24	5391	5209	19.3	260	6.0
25	7406	7194	22	347	9.0
26	12029	11771	27	520	14.8
27	15753	15447	29	663	20
28	17706	17494	45	705	12
29	34380	34020	52	1380	26
B. Six 3-in. elbows and 700 cm straight 3-inch nominal pipe size (R =3.9 cm) iron pipe between pressure taps.					
1	0.75	0.07	0.185	0.0048	0.8
2	1.40	0.10	0.190	0.0095	0.1
3	2.60	0.30	0.21	0.0187	−1.7
4	4.60	0.40	0.24	0.039	−1.2
5	7.95	0.85	0.28	0.077	0
6	13.05	1.35	0.38	0.171	−0.7
7	19.0	3.0	0.47	0.29	0.3
8	29.0	5.0	0.64	0.59	0.7
9	40.0	8.2	0.67	0.82	5
10	48.0	10.0	0.98	1.44	1.4
11	62.0	20.0	1.46	2.36	0.1
12	52.5	11.5	1.17	1.85	0
13	96.0	62.0	2.26	3.0	2.2
14	138.5	109.5	3.3	3.7	3.6
15	178.0	150.0	4.2	4.5	4.3
16	255.0	237.0	9.1	6.3	−1.6
17	430.0	410.0	13.0	10.0	1.0
18	4810	4790	168	129	−0.1
19	8413	8387	215	215	42
20	14020	13980	230	350	15
21	22018	21982	320	440	19
22	48031	47970	440	1070	38
23	69040	68060	540	1600	48

TABLE III. (*Continued*).

Run	p_1	p_2	U	Re	L_e
C. Four 1½-in. Kinney vacuum-tight Bellows valves and 226 cm 1½-in. nominal pipe size (R =2.04 cm) iron pipe between pressure taps.					
1	27	5	0.154	0.25	1.8
2	56.75	12.25	0.25	0.82	3.3
3	35.15	6.85	0.187	0.39	1.7
4	95.55	22.05	0.357	1.95	5.5
5	135.0	40.0	0.50	3.5	6.4
6	165.0	67.0	0.65	4.7	5.9
7	210.0	116.0	0.90	6.2	5.9
8	265.0	165.0	1.10	8.1	6.9
9	318.0	220.0	1.33	9.7	7.5
10	384.5	289.5	1.73	12.1	6.7
11	452.0	358.0	2.16	15	6.0
12	370.0	270.0	1.65	12.2	6.3
13	470.0	380	2.8	19	2.0
14	510.0	410	3.0	22	2.4
15	731.5	618.5	3.9	32	4.2
16	2897	2704	9.1	130	18.5
17	4732	4468	11.2	218	29
18	7574	7226	13.5	350	44
19	11830	11370	15.3	520	66
20	15580	15020	16.1	660	85
21	13420	12980	18.1	580	62
22	26590	26010	23	980	106
23	45980	45020	22	1580	203

near the pipe inlet served to indicate when steady conditions had been attained. All tests were at approximately 26°C. Both air and hydrogen were used in this equipment.

The second series was carried out in a similar way, with provision of higher pumping rates to handle larger pipes. Low pressures were measured with both McLeod and Alphatron gauges. The air flow was measured by an oil displacement device connected upstream from the vacuum reducing valve controlling the flow rate. Air normally passed first through an open stopcock and then through the reducing valve. A large graduated U-tube filled with oil of low vapor pressure was connected to the system between the two valves—the other arm of the U-tube was open to the atmosphere. When the first stopcock was closed air was withdrawn from one arm of the U-tube and the oil level changed at a rate corresponding to the air flow. The actual flow rates were calculated from the observed rate of change of oil level in the U-tube. For very high rates calibrated brass nozzles were employed and air flow rates obtained from observed upstream pressures.

The experimental data on the several straight pipes are summarized in Table II.

RESULTS OF TESTS WITH COPPER AND IRON PIPES

The data of the two series on copper pipes is shown in Fig. 3, where the correction factor F to be applied to Poiseuille's equation (1) is plotted *vs.* the dimensionless group $\mu/p_m R(\rho_1 g_c)^{\frac{1}{2}}$. The dotted line shown is that from Figs. 1 and 2 for the data from the literature on small glass capillaries. This curve also represents Eq. (8) for

$f=0.84$. Not only is the correlation good, but the data evidently agree well with the previous data on small glass tubes.

Figure 4 shows the data on three sizes of large iron pipes. Again the correlation is good, although the solid line representing the data shows a greater curvature than the lines of constant f, and the points near the middle of the curve correspond to values of f greater than unity. The data fall 0 to 30 percent below the curve for glass and copper tubes. The two smaller iron pipes were not new, but were fairly clean, while the largest pipe was dirty and rough. Since the points for all three pipes fall on a single curve no variation in f with pipe conditions is indicated. The solid curve may be taken as a satisfactory empirical correlation of the data on the three sizes of iron pipe.

Maxwell[4] introduced the factor f to allow for the fact that some of the molecules striking the wall are diffusely reflected and some are specularly reflected. The physical meaning of f would suggest that it should not vary with flow conditions, and that it should have the same value in the slip-flow region as in molecular flow. Actually the data on glass capillaries are best fitted by taking f as 0.77 for molecular flow ($X>1.0$) and f as 0.84 for slip-flow ($X<1.0$). Furthermore, the data on iron pipes correspond to a considerable variation of f through the slip-flow region. It seems probable that these results are caused by faults in the theory, rather than to real variations in f. In the slip-flow region the velocity distribution across the pipe would be expected to become flatter as the pressure decreases and internal friction or viscosity becomes less and less significant, and not to maintain its original parabolic shape superimposed on the increasing velocity of slip at the walls. This effect, which is not taken into account in the derivation for slip-flow, would be expected to be greater in the intermediate region of slip-flow, where both internal and wall friction are important. It is significant that the data deviate most from curves of constant f in this region, and that the deviation is in the right direction.

It will be noted that in Table II there appear several points for which F is less than unity. For the copper pipes there are nine tests with X less than 0.01, with an average F of 1.03. For the iron pipes there are fifteen such tests, with an average F of 0.88. This indicates that the observed flow rates in streamline flow are lower than would be calculated by Poiseuille's equation, and suggests that the whole curve of Fig. 4 may be somewhat low. For most purposes this error, if applicable to flow at low pressures, will introduce a small factor of safety in using the solid curve of Fig. 4 as an empirical correlation for iron pipes in design calculations.

TABLE IV. Flow of air in short lengths of straight iron pipe.

Run	p_1	p_2	U	X	K
	A. 23.4 cm of 1-in. pipe ($R=1.33$ cm).				
1	397	27.0	1.57	0.0145	0.058
2	246	4.0	1.16	0.0246	0.070
3	199.5	2.5	1.05	0.0305	0.077
4	155	2.0	0.90	0.039	0.083
5	110	1.0	0.74	0.055	0.091
6	61.1	0.5	0.54	0.10	0.105
7	32	0	0.36	0.192	0.109
8	16.8	0	0.28	0.36	0.109
9	24	0	0.32	0.257	0.109
10	13.5	0	0.27	0.46	0.114
11	6.15	0.05	0.25	1.0	0.12
12	2.90	0	0.23	2.05	0.109
13	2.20	0	0.23	2.8	0.109
14	1.25	0	0.23	4.9	0.101
15	0.88	0	0.22	7.0	0.089
	B. 12.7 cm of 1-in. pipe ($R=1.33$ cm).				
1	334	31.0	1.94	0.017	0.081
2	216.5	3.5	1.49	0.028	0.10
3	148	0	1.29	0.042	0.123
4	119.5	1.5	1.12	0.051	0.129
5	97.5	6.5	1.01	0.066	0.142
6	74	1.0	0.89	0.082	0.150
7	55.5	0.5	0.80	0.11	0.168
8	41.0	0.4	0.74	0.149	0.191
9	32.5	0.5	0.64	0.187	0.190
10	25	0	0.50	0.246	0.168
11	15	0	0.43	0.41	0.173
12	10.6	0.1	0.40	0.58	0.175
13	13	0	0.40	0.47	0.172
14	7.3	0	0.36	0.83	0.168
15	3.1	0	0.38	1.99	0.18
16	1.7	0	0.35	3.42	0.16
17	0.88	0	0.33	7.0	0.13
	C. 6.4 cm. of 1-in. pipe ($R=1.33$ cm).				
1	305	51	2.55	0.017	0.111
2	160	4	1.8	0.037	0.159
3	119	2	1.65	0.05	0.188
4	79	1	1.42	0.077	0.23
5	53	1	1.27	0.114	0.27
6	35.5	0.5	1.09	0.171	0.30
7	28	0	0.95	0.22	0.31
8	18	0	0.79	0.34	0.31
9	10.8	0	0.70	0.57	0.31
10	5.0	0	0.80	1.23	0.38
11	3.2	0	0.60	1.92	0.28
12	1.6	0	0.57	3.8	0.26
13	0.58	0	0.62	10.6	0.27
14	0.40	0	0.55	15.4	0.24

COMPARISON WITH THE KNUDSEN EQUATION

The Knudsen semi-empirical equation[12,13] may be written

$$Q=\frac{\pi}{16}\frac{g_c}{\mu}\frac{R^4}{L}(p_1^2-p_2^2)\times\left[1+8.52X\left(\frac{X+2.00}{X+2.47}\right)\right] \quad (15)$$

or,

$$F=1+8.52X\left(\frac{X+2.00}{X+2.47}\right). \quad (16)$$

This is found to fit the curves of Figs. 1–3 within five percent or less, and so represents the data on glass and copper tubes essentially as well as the empirical lines drawn to represent the data. The Knudsen equation does not, however, fit the data on iron pipe, as indicated by Fig. 4, which includes a dotted curve representing both the copper and glass tube data and Eq. (15).

RESULTS WITH ELBOWS AND VALVES

Table III gives the data obtained by the same technique, using lines in which were inserted: (a) eight 1½-inch elbows, connected by short (6–8 inch) sections of 1½-inch nominal diameter iron pipe; (b) six 3-inch elbows, connected by 36-inch sections of 3-inch nominal diameter iron pipe; and (c) four 1½-inch Kinney vacuum-tight bellows valves connected by sections of 1½-inch iron pipe. The last two columns in each case give values of the Reynolds number, Re, and the length of straight pipe, L_e, expressed in pipe diameters, corresponding to the resistance of one valve or elbow. Values of L_e were obtained by comparing the observed values of S with those calculated from (13) for the actual total length of pipe, using values of F from the solid line of Fig. 4.

The results indicate that neither the valves nor the elbows offer appreciable resistance at low pressures. For Reynolds numbers below 100, L_e can be taken as unity for the elbows, and may be assigned the nominal value of five for the Kinney valve. As Re increases above 100, L_e for the valves becomes appreciable, approaching the accepted value[14,15] of about 30 for turbulent flow ($Re>2100$). In the case of the valve L_e is considerably greater, increasing from 5 at a Re of about 50 to greater than 200 at the critical value of Re. The data for the elbows are erratic in the range of Re from 100 to 2000, for which reliable correlations exist.[15] The data show clearly, however, that the resistance of elbows may be neglected at low pressures, where Re is less than 100.

RESULTS WITH SHORT TUBES

Data were obtained on flow through three short lengths of 1-inch iron pipe connecting two large domes. Air was supplied to one dome through a controlled leak, and the pressures in both domes measured. The data obtained are given in Table IV.

Values of S were calculated from Eq. (13) and the solid line of Fig. 4. In the molecular flow region the experimental data are represented by Eq. (11) and the calculated values of S by Eq. (9), using f of approximately 0.9. Hence K is obtained as

$$\frac{\pi}{2}\left(\frac{2}{f}-1\right)\frac{R}{L}$$

times the ratio of experimental to calculated values of S. The values of K obtained for the molecular flow region average 0.11, 0.16, and 0.29 in the three series, which compare well with the values 0.12, 0.20, and 0.31 given by Clausing[7] for the three values of L/R. The Clausing factor K is not properly applicable outside the molecular flow region, as indicated by the way the tabulated values vary and decrease very considerably at higher pressures.

The combined pressure drop due to the contraction and the flow through the tube may be obtained by adding the pressure drops calculated from Eq. (10) for the entrance and Eq. (13) for flow in the straight pipe. If p_1 represents the pressure before contraction, p_1' the pressure in

[12] S. Dushman, J. Frank. Inst. 211, 689 (1931).
[13] L. B. Loeb, *The Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1934), second edition, Chapter VII.
[14] J. H. Perry, Editor, *Chemical Engineers' Handbook* (McGraw-Hill Book Company, Inc., New York, 1941), second edition, p. 822.
[15] R. E. Wilson, W. H. McAdams, and M. Seltzer, Ind. Eng. Chem. 14, 105 (1922).

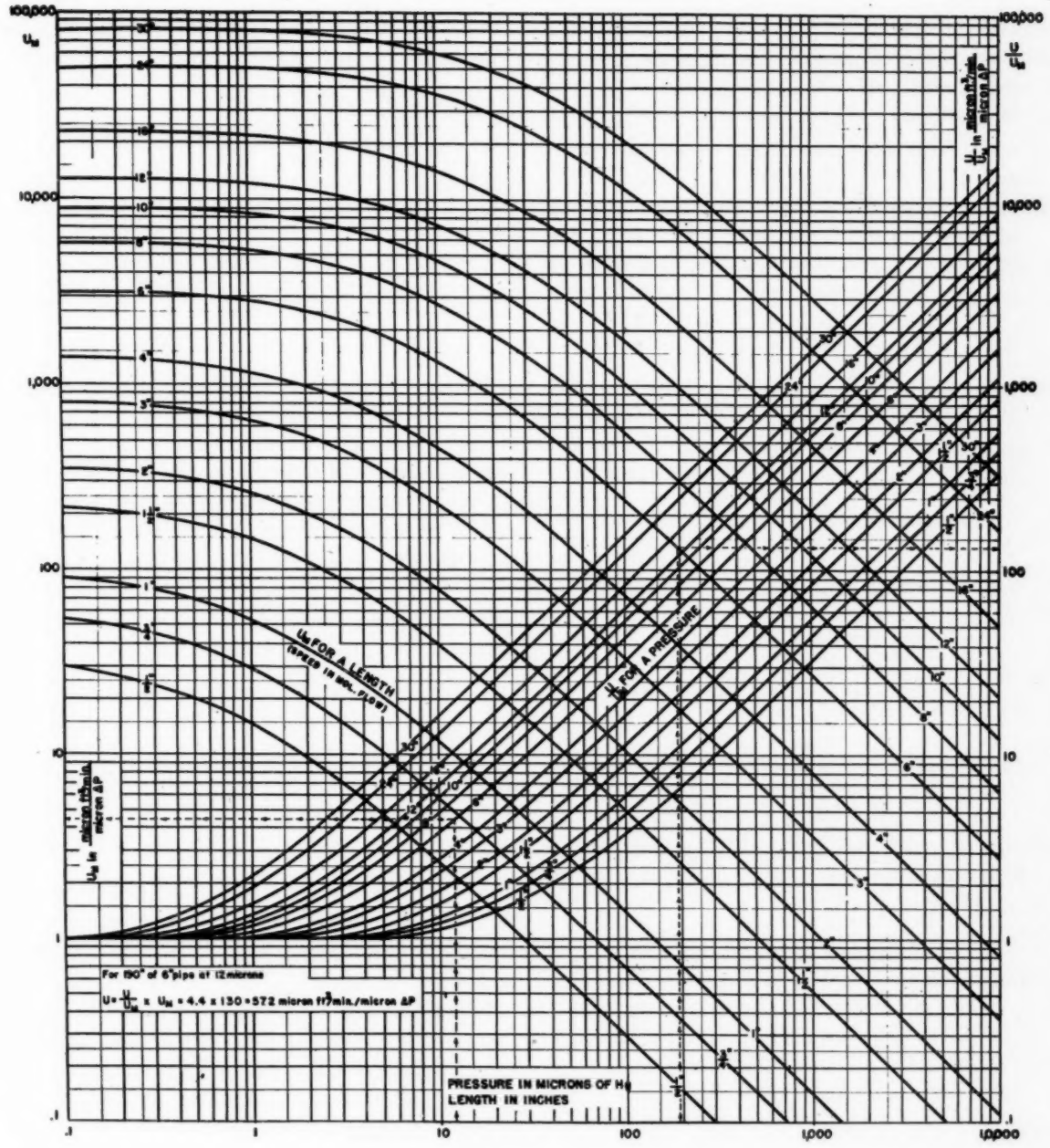


FIG. 5.

the tube at the inlet end, and p_2 the pressure at the downstream end, then

$$\frac{1}{S}=\frac{p_1-p_2}{Q}=\frac{p_1-p_1'}{Q}+\frac{p_1'-p_2}{Q}$$

$$=B\left(\frac{2\rho_1}{\mu g_c R^4}\right)^{\frac{1}{2}}+\frac{8\mu L}{F\pi R^4 p_m g_c}. \quad (17)$$

In order to compare this equation with the data of Table IV, it was assumed that a correction term B would necessarily be introduced as a multiplying factor before the first term. A value of p_1' was arrived at by trial and error calculation such that $p_1'-p_2$ agreed with the pressure drop calculated from Eq. (13) using F obtained using Fig. 4 and $p_m=(p_1'-p_2)/2$. The values of B

obtained in this way varied somewhat due to the fact that the first term of the right side of Eq. (17) was a small difference of two larger numbers, but the average value was about 0.5. It is suggested, therefore, that Eq. (17) be used with a value of B of 0.5 to allow for entrance and exit losses in flow through short tubes. Values of U calculated in this way for the tests of Table IV showed an average deviation of less than 10 percent from the measured values.

As an example, suppose it is desired to calculate S for flow of air at 26°C through a tube 12.7 cm long and 1.33 cm radius when the pressures outside the tube upstream and downstream are 97.5 and 6.5 microns, respectively. Using microns-c.g.s. units, with $\mu=1860\times10^{-7}$ poises, and

$$\rho_1=\frac{29}{22{,}400}\times\frac{273}{299\times760{,}000}=1.56\times10^{-9},$$

then

$$X=\frac{\mu}{p_m R(\rho_1 g_c)^{\frac{1}{2}}}$$

$$=\frac{1860\times10^{-7}}{p_m\times1.33(1.33\times1.56\times10^{-9})^{\frac{1}{2}}}=\frac{3.05}{p_m}.$$

Assume $p_1'=90$ microns, then $p_m=48.2$, $X=0.064$, and from Fig. 4, $F=1.21$. Substituting in Eq. (17),

$$\frac{1}{S}=0.5\left(\frac{2\times1.56\times10^{-9}}{3.14\times1.33\times1.33^4}\right)^{\frac{1}{2}}$$

$$+\frac{8\times1860\times10^{-7}\times12.7}{1.21\times3.14\times1.33^4\times48.2\times1.33}$$

$$=0.77\times10^{-5}+2.47\times10^{-5}=3.24\times10^{-5}.$$

These two terms should be proportional to the two parts of the total pressure drop, so

$$p_1'=6.5+\frac{2.47}{3.24}(97.5-6.5)=76 \text{ microns},$$

which does not check the assumed value of 90. Repeating the calculation until a check is obtained, it is found that $p_1'=77.5$, $S=29{,}600$ micron-cu. cm per sec. per micron, or $U=1.01$ micron-cu. ft. per sec. per micron. This checks the observed value obtained in Run B-5 (Table IV).

SUMMARY

Data on flow of gases in round pipes at low pressures are well correlated by the equation,

$$S=\frac{Q}{(p_1-p_2)}=\frac{F\pi R^4 p_m g_c}{8\mu L}, \qquad (13)$$

where F is related empirically to the dimensionless group X by Figs. 1–3 for glass and copper, and Fig. 4 (solid line) for iron pipes. The pressure p_m is the arithmetic mean of the upstream and downstream pressures, p_1 and p_2. Any consistent set of units may be employed, such as the four given in Table I, and the correlation may be used for streamline flow, slip flow, or molecular flow.

In c.g.s.-micron units the following simplifications may be used for air or water vapor at 26°C.

For air, $X=\dfrac{4.10}{p_m R}$; $S=2910\dfrac{FR^4p_m}{L}$.

For water vapor, $X=\dfrac{2.66}{p_m R}$; $S=5500\dfrac{FR^4p_m}{L}$.

(In units of cu. ft., microns, and minutes, with R and L in inches, the constants become 1.61, 101, 1.05, and 191 in place of 4.10, 2910, 2.66, and 5500, respectively.)

For Reynolds numbers below 100 (most vacuum engineering) the resistance of 1½- and 3-inch elbows may be neglected; a 1½-inch Kinney bellows valve has a resistance equivalent to about five pipe diameters.

For flow through short tubes involving both contraction and expansion, the expansion loss is neglected and the flow or pressure drop calculated from Eq. (17), using the method of calculation illustrated by a numerical example in the preceding section.

As an example of the method of calculation, assume that it is desired to determine the "speed" of a standard 2-inch iron pipe with upstream and downstream pressures of 30 and 2.0 microns, respectively, at points 10 feet apart.

The gas is water vapor at 26°C.

$$p_m = \frac{30.0+2.0}{2} = 16.0 \text{ microns},$$

$$R = 2.07 \times 2.54/2 = 2.63 \text{ cm},$$

$$L = 10 \times 30.48 = 304.8 \text{ cm},$$

$$X = \frac{2.66}{16.0 \times 2.63} = 0.063,$$

and from Fig. 4, $F = 1.41$,

$$S = 5500 \frac{1.21 \times 2.63^4 \times 16}{304.8}$$

$$= 16{,}500 \text{ micron-cu. cm/(sec.)(micron)},$$

$$U = 16{,}500/28{,}400$$

$$= 0.58 \text{ micron-cu. ft./(sec.)(micron)}.$$

The Reynolds number is $2Q\rho_1/\pi\mu$, and since

$$\rho_1 = \frac{18}{82.07 \times 299 \times 760{,}000} = 9.65 \times 10^{-10} \text{ g/cu. cm}$$

and $\mu = 950 \times 10^{-7}$ g/(sec.) (cm), then

$$Re = \frac{2 \times 16{,}500(30-2) \times 9.65 \times 10^{-10}}{3.14 \times 950 \times 10^{-7}} = 2.99.$$

Since this is well below 100, it may be assumed that the resistance of elbows will be negligible.

Equation (13) and the graphical correlations of F and X may be combined in the form of a single graph or alignment chart for engineering use. Figure 5 is an example of a useful chart prepared in this way, based on a constant f of 0.90. In using the chart, the speed in molecular flow is first obtained from the ordinate of the appropriate diameter curve, using length in inches as abscissa and the curves sloping downward to the right. The ratio of actual speed to speed in molecular flow is then obtained as the ordinate using the arithmetic mean pressure p_m in microns as abscissa, and one of the family of lines sloping up to the right. The product of the two figures is U, the admittance of the pipe, and may be used in the usual reciprocal speed formula. The curves hold for all combinations of pressure and pipe size except for very short lengths in the streamline region, where the Clausing correction to U_m loses meaning. Fortunately, however, speed evaluations for short pipe at high pressure is seldom critical.

TABLE OF NOMENCLATURE

The units quoted are in the micron-c.g.s. system except where otherwise noted; see the paragraph "Units" regarding use of other systems.

B = correction factor to be applied to the molecular effusion equation to give pressure drop because of contraction.
f = fraction of molecules striking wall which is diffusely reflected.
g_c = factor for converting microns Hg weight to microns Hg mass, = 1.33.
k = coefficient of external friction.
K = Clausing coefficient, in Eq. (11).
L = tube length, cm.
L_e = length of straight pipe giving same pressure drop as valve or elbow diameters.
M = molecular weight of gas.
p_1 = upstream pressure, microns Hg.
p_2 = downstream pressure, microns Hg.
p_m = arithmetic mean pressure = $(p_1+p_2)/2$.
Q = gas flow rate, micron-cu. cm/second.
R = tube radius, cm.
R_G = gas law constant = 6.24×10^{-7}.
S = rate of flow per unit pressure drop, micron-cu. cm/(sec.) (micron).
T = temperature, °K.
U = rate of gas flow per unit pressure drop = $Q/\Delta p$, micron-cu. ft./(sec.) (micron).
u_0 = velocity of slip at wall, cm/sec.
U_m = U in molecular flow.
$X = \mu/p_m R(\rho_1 g_c)^{\frac{1}{2}}$.
λ = mean free path of gas molecules, cm.
λ_m = mean free path of gas molecules at mean pressure p_m, cm.
ρ_1 = gas density at one micron Hg, g/cu. cm.
μ = gas viscosity at atmospheric pressure, poises.

Absorption of Microwaves by Gases. II

JOHN E. WALTER* AND W. D. HERSHBERGER
RCA Laboratories, Princeton, New Jersey
(Received May 2, 1946)

The absorption coefficients and dielectric constants of sixteen gases have been measured at the two wave-lengths $\lambda=1.24$ cm and $\lambda=3.18$ cm. The gases are H_2S, COS, $(CH_3)_2O$, C_2H_4O, C_2H_5Cl, SO_2, NH_3, six halogenated methanes and three amines. Certain improvements in technique are described; these improvements permit detection of absorption coefficients as small as 0.2×10^{-4} cm^{-1} and measurement of larger coefficients with an accuracy of ±5 percent. The measured dielectric constants at these wave-lengths are essentially equal to the static values. A quantitative interpretation of the absorption coefficients in terms of the known structure and spectra of the individual molecules is given. The theory indicates that all non-planar molecules which possess a permanent dipole moment should show appreciable absorption in the microwave region.

INTRODUCTION

RESULTS of an investigation on the absorption of microwaves by gases were given in an earlier paper.[1] Measurements on absorption are taken by noting the power loss suffered by a microwave transmitted through a rectangular wave guide operating in the TE_{10} mode. Dielectric constants are measured by comparing the wave-length in the guide when evacuated with the wave-length after the gas under study is admitted to the guide. Measurements are reported at two wave-lengths: 3.18 cm and 1.24 cm. The experimental values of the absorption coefficients α (in cm^{-1}) are given in Table I; the experimental values of the dielectric constants are listed in Table II. The purpose of the present paper is to report improvements in technique, to give the theory underlying the observed absorption, and to present new data and analyses of these data.

TABLE I. Observed absorption coefficients.

	$\alpha\times10^4$ ($\lambda=3.18$)	$\alpha\times10^4$ ($\lambda=1.24$)
NH_3	15.5	78.0
CH_3F	7.6	10.0
CH_3Cl	5.5	8.0
CH_3Br	4.2	6.2
C_2H_5Cl	8.7	15.0
$CHFCl_2$	5.2	10.6
CHF_2Cl	7.2	12.0
CF_2Cl_2	0.3	3.1
H_2S	~0.0	~0.3
SO_2	1.1	5.0
COS	—	0.8
$(CH_3)_2O$	0.7	3.9
C_2H_4O	1.0	6.8
CH_3NH_2	2.1	10.5
$(CH_3)_2NH$	—	7.0
$C_2H_5NH_2$	3.1	10.2

* Present Address: Palmer Physical Laboratory, Princeton University, Princeton, New Jersey.

[1] W. D. Hershberger, J. App. Phys. **17**, 495 (1946).

TECHNIQUES AND ACCURACY OF MEASUREMENT

One factor which has limited the accuracy with which small absorption coefficients may be measured is the standing wave in the test section of guide. This effect was troublesome in the work reported earlier but has now been eliminated by placing a wedge of lossy solid material in this section near the input window. The loss introduced by the wedge is 10 db and, since it is not inserted until the standing wave has been minimized by the usual tuning adjustments, no measurable standing wave now exists when measurements are taken. The procedure for tuning is first to fill the 30-foot test section with ammonia which effectively isolates the receiver

TABLE II. Dielectric constants.

	$\delta_0\times10^3$ (calculated)	$\delta\times10^3$ ($\lambda=3.18$)	$\delta\times10^3$ ($\lambda=1.24$)
NH_3	6.3	5.3	5.5
CH_3F	9.2	8.1	8.4
CH_3Cl	10.3	9.4	9.9
CH_3Br	9.7	9.5	10.0
C_2H_5Cl	12.7	11.1	11.6
$CHFCl_2$	6.2	4.9	5.2
CHF_2Cl	6.7	5.2	5.7
CF_2Cl_2	3.1	2.9	3.1
H_2S	3.6	3.0	3.5
SO_2	7.9	7.5	8.6
COS	2.6	—	2.6
$(CH_3)_2O$	6.2	5.5	6.1
C_2H_4O	10.4	10.2	10.7
CH_3NH_2	5.3	4.8	5.3
$(CH_3)_2NH$	4.4	—	4.0
$C_2H_5NH_2$	6.0	5.3	4.5

from the generator and then adjust the tuner which is adjacent to the input window for a minimum standing wave. Next, the ammonia is removed and the tuner adjacent to the output window is adjusted for a minimum standing wave. As a result of this adjustment receiver power very nearly attains its maximum value. Finally, the 10-db wedge is placed inside of the test section of guide and the operation of the system is relatively insensitive to changes in the frequency of the generator. The result is that an absorption coefficient as small as 0.2×10^{-4} neper per cm may be measured, but of course the accuracy in this range is low. Thus if the received power falls to 98 percent of its initial value on introducing a gas into a 30-foot section of guide the absorption coefficient is 0.18×10^{-4} neper per cm, while if the power drops to 97 percent of its initial value, the coefficient is 0.27×10^{-4} neper per cm. When a gas with large absorption is used, the fact that absorption vanishes as the pressure is reduced is of great aid in following the course of a pressure *vs.* absorption curve.

When the absorption coefficient is of the order of 10^{-3} neper per cm or 40 times the minimum measurable absorption, the consistency of the results obtained is such that the accuracy of the absorption coefficients reported is estimated to be somewhat better than 5 percent. Approximately the same accuracy is realized in the measurement of δ, which is the amount by which the dielectric constant exceeds unity.

Figure 1 represents the experimental points obtained in six absorption runs with methyl chloride. These readings lie in the range of pressure and absorption for which accuracy is high, using the 30-foot section of guide.

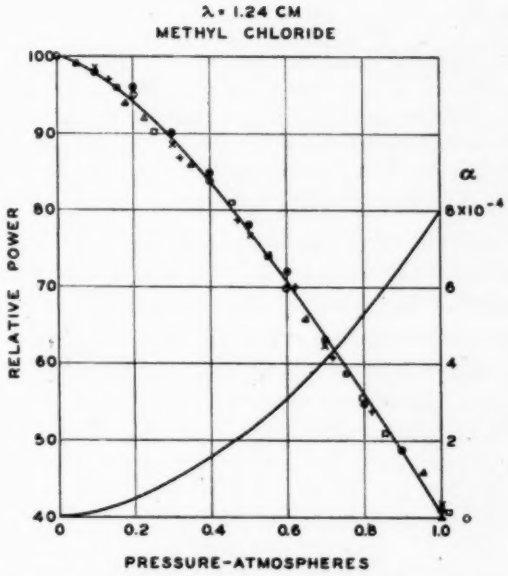


FIG. 1. Absorption of methyl chloride at $\lambda=1.24$ cm.

GENERAL THEORY

The essential parameters which characterize an absorption line are ν_0, the frequency of the center of the line, and the half-width $\Delta\nu$, which is the difference between ν_0 and the frequency at which the absorption has fallen to one-half of the absorption at the center of the line. In the microwave region the observed broadening of an absorption line is to be attributed to collision broadening; Doppler broadening and natural line breadth are negligible at these frequencies. Since at one atmosphere $\Delta\nu$ is found to be of the same order of magnitude as ν_0 if ν_0 is in the microwave region, the exact shape of the absorption line becomes a matter of considerable interest. This problem of the shape of collision-broadened lines has been re-examined by Van Vleck and Weisskopf,[2] who have shown that the classical formula of Lorentz requires revision. The fundamental difference between the assumptions made in the two theories, in applying them to absorption by molecules possessing a permanent dipole moment, is the following: Lorentz assumed that after each collision all orientations of the dipole with respect to the direction of the electric field are equally probable; Van Vleck and Weisskopf assumed that the probability of a given orientation is given by the corresponding Boltzmann factor. The absorption coefficient α_ν for electromagnetic radiation of frequency ν is calculated by the latter authors to be

$$\alpha_\nu=\frac{8\pi^3\nu N}{3hc}\cdot\frac{\sum_j\sum_i|\mu_{ij}|^2 f(\nu_{ij}\nu)}{\sum_j e^{-W_j/kT}}e^{-W_j/kT}, \qquad (1)$$

$$f(\nu_{ij}\nu)=\frac{1}{\pi}\frac{\nu}{\nu_{ij}}\left\{\frac{\Delta\nu}{(\nu_{ij}-\nu)^2+\Delta\nu^2}+\frac{\Delta\nu}{(\nu_{ij}+\nu)^2+\Delta\nu^2}\right\}, \qquad (2)$$

[2] J. H. Van Vleck and V. F. Weisskopf, Rev. Mod. Phys. 17, 227 (1945).

where

N = number of molecules per cc,
μ_{ij} = electric dipole moment matrix element for the two states i and j which have a frequency separation ν_{ij},
W_j = energy of jth state, and
$\Delta\nu$ = half-width of a line $= 1/2\pi\tau$, where τ is the mean time between molecular collisions.

The other symbols have their usual significance. The summations are to be extended over all the states of the molecule. Equations (1) and (2) differ from the Lorentz result by the inclusion of the factor ν/ν_{ij} in (2) and by the change in the sign of the second term in (2). For a molecule with large moments of inertia and a permanent dipole moment μ_0, these equations reduce to the familiar Debye formula

$$\epsilon_\nu'' = \frac{\alpha_\nu c}{2\pi\nu} = \frac{4\pi N\mu_0^2}{3kT}\,\frac{\nu\Delta\nu}{\nu^2+\Delta\nu^2} \tag{3}$$

for the complex part of the dielectric constant.

For our purposes it is convenient to specialize Eqs. (1) and (2) so that they apply directly to the ammonia problem. NH_3 is a symmetrical top molecule with rotational energy levels specified by the quantum numbers J and K.[3] Each rotational level is split into a close doublet as a result of the tunneling effect; the two components of the doublet are specified by the designation $+$ and $-$. The selection rules for rotational transitions are $\Delta J=0, \pm1$; $\Delta K=0$; $+\leftrightarrow-$. The "inversion spectrum" frequency corresponding to $\Delta J=0$, $\Delta K=0$; $+\leftrightarrow-$ is about 0.8 cm^{-1} while the lowest frequency in the rotation spectrum corresponding to $\Delta J=\pm1$ is about 20 cm^{-1}. For $\nu<1.0$ cm^{-1} only the absorption associated with the inversion spectrum need be considered, and Eq. (1) can be reduced to[2]

$$\alpha_\nu = \frac{8\pi^3\nu N}{6ckT}\,\frac{\sum_j\sum_i|\mu_{ij}|^2 f(\nu_{ij},\nu)\nu_{ij}e^{-W_j/kT}}{\sum_j e^{-W_j/kT}}, \tag{4}$$

where j and i now refer to the two components of a given doublet. The reduction follows from the relation $h\nu_{ij}\ll kT$. Aside from the negligible difference in the Boltzmann factor, we obtain the same contribution to the sum in the numerator from the transition $j\rightarrow i$ as from the transition $i\rightarrow j$. If $W(J,K)$ is the mean energy of a given doublet, and $g(J,K)$ is the total weight of the rotational level with quantum numbers J and K, Eq. (4) becomes

$$\alpha_\nu = \frac{4\pi^2N\nu^2}{3ckT}\,\frac{\sum_{J,K} g(J,K)|\mu(J,K)|^2 f'[\nu(J,K),\nu]e^{-W(J,K)/kT}}{\sum_{J,K} g(J,K)e^{-W(J,K)/kT}}, \tag{5}$$

where

$$f'[\nu(J,K),\nu] = \frac{\Delta\nu}{[\nu(J,K)-\nu]^2+\Delta\nu^2} + \frac{\Delta\nu}{[\nu(J,K)+\nu]^2+\Delta\nu^2}, \tag{6}$$

and where the sums over the rotational levels in the excited vibrational states have been omitted. $\mu(J,K)$ is the electric moment matrix element for the transition $\Delta J=0$, $\Delta K=0$, $+\leftrightarrow-$ and is equal to[4]

$$|\mu(J,K)|^2 = \frac{K^2}{J(J+1)}\mu_0^2, \tag{7}$$

where μ_0 is the permanent dipole moment of the molecule. For an accidentally symmetrical top the weights $g(J,K)$ are $(2J+1)$ for $K=0$, and $2(2J+1)$ for $K\neq0$. For NH_3 or CH_3Cl the presence of the threefold symmetry axis and the nuclear spin of the hydrogen atoms introduce the additional factor 4 for K divisible by 3 and the factor 2 for K not divisible by 3.[5] Since both numerator and denominator of (5) involve the same set of weights, only the relative values are important. Factors such as the spin of N or Cl, which would change the absolute weights, will not affect the relative weights. In the sums we can therefore use the weights

$g(J,K) = (2J+1)$ times $4,4,4,8,4,4,8,4\cdots$

[3] G. Herzberg, *Infra-Red and Raman Spectra* (D. Van Nostrand Company, New York, 1945), pp. 26–34.
[4] Reference 3, p. 422.
[5] Reference 3, p. 28.

for

$$K=0, 1, 2, 3, 4, 5, 6, 7\cdots. \quad (8)$$

If these weights are used, the sum in the denominator, which we denote by Q, becomes (in the high temperature approximation)

$$Q=8/3(Q_{\text{class}})=8/3(1.027)[T^3/B^2A]^{\frac{1}{2}}, \quad (9)$$

where Q_{class} is the classical rotational partition function for a symmetrical top with rotational constants B and A, and T is the absolute temperature. The energy levels, in wave numbers, are

$$W(J, K)=BJ(J+1)+(A-B)K^2. \quad (10)$$

If $\nu(J, K)=\nu_0$ independently of J and K, the absorption coefficient is given by

$$\alpha_\nu=\beta N\nu^2\left[\frac{\Delta\nu}{(\nu_0-\nu)^2+\Delta\nu^2}+\frac{\Delta\nu}{(\nu_0+\nu)^2+\Delta\nu^2}\right], \quad (11)$$

where β is a constant which can be calculated from the sums in (5). We also note that

$$\int_0^\infty f'[\nu(J, K), \nu]d\nu=\pi, \quad (12)$$

independently of $\nu(J, K)$ and $\Delta\nu$. Therefore, from (12) and (5), we have, as a measure of the total absorption arising from this inversion spectrum

$$\int_0^\infty \frac{\alpha_\nu d\nu}{\nu^2}$$

$$=\frac{4\pi^3N}{3ckT}\frac{\mu_0^2}{Q}\sum_{J,K}g(J, K)\frac{K^2}{J(J+1)}e^{-W(J,K)/kT} \quad (13)$$

independently of any assumptions as to the detailed positions or breadths of the individual lines.

Since our measurements are limited to two frequencies, they do not give any direct information on the half-widths $\Delta\nu$ of the absorption lines. From the kinetic theory of gases and the relation $\Delta\nu=1/2\pi\tau$ one would expect to find $\Delta\nu\sim0.05$ cm^{-1} or less at room temperature and atmospheric pressure. Adel and Barker[6] find $\Delta\nu=0.067$ cm^{-1} for zero slit width for the lines in N_2O, which is a linear molecule with a very small dipole moment. For molecules with a large dipole moment much larger values of $\Delta\nu$ have been reported; for example, Cornell[7] reports $\Delta\nu\sim0.3$ for H_2O and larger values for NH_3 and HCN (not corrected for slit width). These larger values are to be expected in dipole molecules, since a "collision," as the word is applied to line broadening, means that the molecules have come sufficiently close to cause a perturbation of the energy levels, not that they have collided in a billiard-sense. The "optical" collision diameters as computed from observed values of $\Delta\nu$ are usually several times larger than the kinetic theory diameters.

[6] A. Adel and E. F. Barker, Rev. Mod. Phys. **16**, 236 (1944).

AMMONIA

Although the absorption of microwaves by NH_3 has been known for a long time and the general features understood,[8] the details of the structure of the line are not clear. Let us assume for the moment that the $\nu(J, K)$'s are all essentially equal, so that the absorption is given by (11). Since $\Delta\nu$ is inversely proportional to the mean time τ between collisions, and since τ is inversely proportional to the pressure P, $\Delta\nu\sim P$. Also, $N\sim P$. For the non-resonant case, with ν_0 very different from ν, Eq. (11) shows that $\alpha_\nu\sim P^2$ as long as $(\nu_0-\nu)\gg\Delta\nu$. For the resonant case, with $\nu_0\cong\nu$, α_ν should rise very rapidly with increasing pressure and become relatively independent of P when $\Delta\nu\gg(\nu_0-\nu)$. These conclusions are in qualitative agreement with the observed behavior of the pressure dependence of the NH_3 absorption (Fig. 2); the curve for $\nu=0.81$ cm^{-1} ($\lambda=1.24$ cm) is of the type corresponding to $\nu\cong\nu_0$, that for $\nu=0.32$ cm^{-1} ($\lambda=3.18$ cm) is of the type corresponding to $\nu\ll\nu_0$. The situation is less favorable if looked at from a quantitative viewpoint. The NH_3 curve for $\nu=0.81$ rises linearly with pressure above 0.1 atmos., and this rise continues to at least 2.5 atmos. The details of the curve cannot be accounted for by Eq. (11) if $\Delta\nu$ is assumed to be proportional to P. Any other variation of $\Delta\nu$ with P would have no theoretical justification at present. If we assume $\nu_0\cong0.80$ cm^{-1}, α_ν should decrease with frequency more rapidly than ν^2 in the region $\nu<0.80$ cm^{-1}. This is in serious disagreement with the observed decrease. Actually, the $\nu(J, K)$'s are not identical.

[7] S. D. Cornell, Phys. Rev. **51**, 739 (1937).

[8] C. E. Cleeton and N. H. Williams, Phys. Rev. **45**, 234 (1934). A complete discussion of the ammonia spectrum and references to the original literature may be found in reference 3, pp. 33, 221, 257, 416.

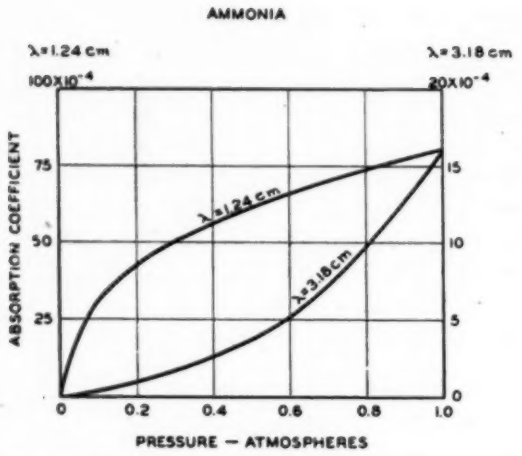


FIG. 2. Absorption of ammonia employing 5-foot guide.

The centrifugal distortion of the molecule caused by rotation produces a change in the effective barrier height and thus a change in the doublet separation. This dependence of the doublet separation on J and K has been calculated by Sheng, Barker, and Dennison,[9] who obtain the result

$$\nu(J, K)-\nu_0=-0.0011(J^2+J)+0.0016K^2, \quad (14)$$

where ν_0 is the doublet separation in the non-rotating molecule. If the Boltzmann distribution is taken into account, one finds that there are lines of appreciable intensity spread over a region about 0.06 cm^{-1} wide (at zero pressure). If the ammonia inversion spectrum consists of a series of lines starting at $\nu\sim 0.80$ cm^{-1} and extending to considerably lower frequencies, the observed features of the absorption curve are more readily understood. This distribution of frequencies would increase the absorption at $\nu=0.32$ cm^{-1} relative to that at $\nu=0.81$ cm^{-1} above that expected for a single line near $\nu=0.80$ cm^{-1}, and would also produce a pressure variation similar to that observed at $\nu=0.81$ cm^{-1}. Such a distribution would also explain the apparent discrepancy between the doublet separation as obtained from the pressure curves ($\nu_0\cong 0.80$ cm^{-1}) and from the far infra-red spectra ($\nu_0\cong 0.66$ cm^{-1}) as observed by Wright and Randall.[10] It would seem, however, that a spread greater than that given by Eq. (14) would be required. Van Vleck and Weisshopf[2] have pointed out that a variation of $\Delta\nu$ with J and K would also tend to increase the absorption at the edge of the line relative to the absorption at the center. It appears that a quantitative description of the NH_3 absorption must be postponed until an investigation at low pressures reveals further information about the $\nu(J, K)$'s.

Application of Eq. (13) to NH_3 yields

$$\int_0^\infty \frac{\alpha_\nu d\nu}{\nu^2}=74\times 10^{-14} \text{ (theoretical)}. \quad (15)$$

(The NH_3 constants are $A=6.31$, $B=9.94$, $\mu_0=1.49\times 10^{-18}$. The ratio $\sum/Q$, where $\sum$ represents the sum in (13), is equal to 0.39 for NH_3 at $T=300°K$.) From the data of Cleeton and Williams[8]

$$\int_0^\infty \frac{\alpha_\nu d\nu}{\nu^2}\cong 35\times 10^{-14} \text{ (experimental)}. \quad (16)$$

The origin of this discrepancy is not clear. Comparisons of this type often lead to even more serious disagreements,[11] but these are usually attributed to errors arising from finite slit width. Foley and Randall[12] have made a similar comparison on the line $J=14\rightarrow J=15$ of NH_3 at a pres-

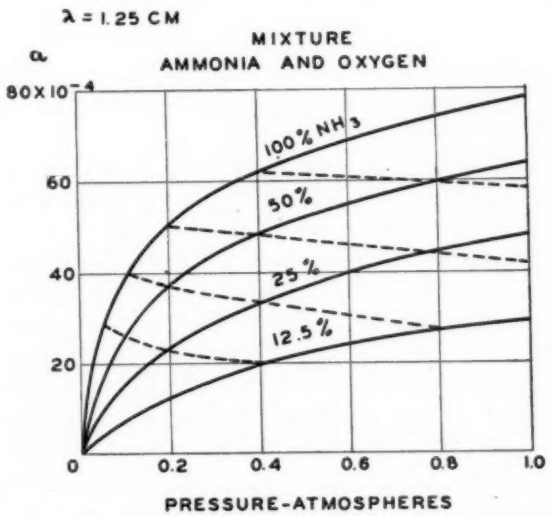


FIG. 3. Absorption of mixtures of ammonia and oxygen.

[9] H. Y. Sheng, E. F. Barker, and D. M. Dennison, Phys. Rev. 69, 786 (1941).
[10] N. Wright and H. M. Randall, Phys. Rev. 44, 391 (1933).
[11] C. Shaefer and F. Matossi, *Das Ultrarote Spectrum* (Verlagsbuchhandlung, Julius Springer, Berlin, 1930).
[12] H. M. Foley and H. M. Randall, Phys. Rev. 59, 171 (1941).

sure of 20 cm and have found the experimental value to be only about 20 percent lower than the theoretical value. It would be of interest to see whether or not the above discrepancy persists at low pressures.

Figure 3 shows the effect of the admixture of oxygen on the ammonia absorption. Almost identical curves are obtained from NH_3-CH_4 and NH_3-H_2 mixtures. The dotted curves connect points of equal NH_3 partial pressure; if these lines were horizontal it would mean that the NH_3 absorption is a function of the NH_3 partial pressure only, since the absorption of O_2 is negligible. The addition of O_2 to NH_3 changes the line breadth $\Delta\nu$ in the latter molecule and so effects the absorption. We can make a rough estimate of the relative collision cross sections for NH_3-O_2 and NH_3-NH_3 collisions as follows. If the pressure of NH_3 is greater than 0.1 atmos., the absorption is, according to Eq. (11), essentially proportional to $1/\Delta\nu$ if the concentration of NH_3 is constant. If we follow the dotted line which starts at $\alpha\times10^4=40$, we find that $\Delta\nu$ increases by a factor $(1.00+0.48)$ as the total pressure is increased from 0.1 atmos. to 0.8 atmos. by the addition of oxygen. A corresponding increase in pressure caused by the addition of NH_3 should increase $\Delta\nu$ by a factor $(1.00+7.00)$. The relative cross sections for the two-collision process are thus in the ratio 14.5:1, and the relative collision diameters about in the ratio 4:1. Since optical collision diameters are considerably larger than kinetic theory diameters, we see that, roughly at least, the NH_3-O_2 collision diameter is close to the kinetic theory value.

METHYL HALIDES

The methyl halides CH_3F, CH_3Cl, CH_3Br all show the type of variation $\alpha_\nu\sim P^2$ which is characteristic of absorption at a frequency far removed from the resonant frequency (Fig. 4). In this figure the curves for CH_3Cl are not shown. For this gas, at a given wave-length, the curve absorption coefficient *vs.* pressure lies between the corresponding curves for CH_3F and CH_3Br. The absolute magnitude of the absorption, the variation of α_ν with ν, and the similarity of the absorption in the three gases are all inconsistent with the assumption that the observed absorption arises from rotational transitions. Comparing the

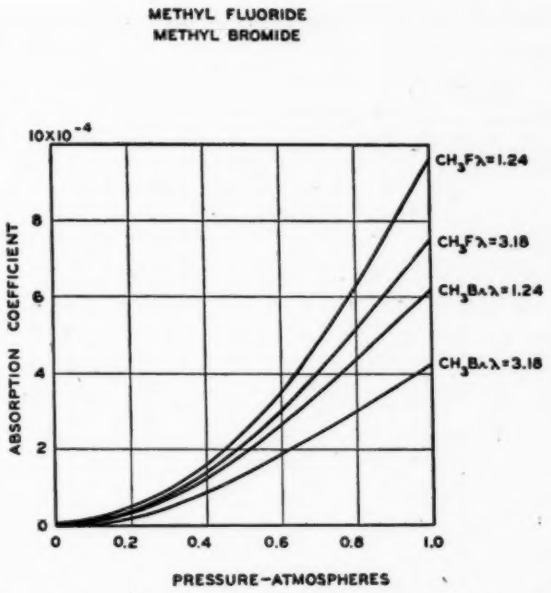


FIG. 4. Absorption of methyl fluoride and methyl bromide.

observed absorption at $\nu=0.81$ cm^{-1} and $\nu=0.32$ cm^{-1} with Eq. (11), we conclude immediately that ν_0 is very small.

From a formal viewpoint, the energy levels of CH_3X are exactly the same as those of NH_3, and the selection rules are identical. Each rotational level characterized by the quantum numbers J and K is double, the two components being designated by the symbols + and −. The separation of this doublet depends on the height of the potential barrier which must be overcome if the molecule is to be turned into its mirror image. In NH_3 this barrier (for the passage of N through the plane of the H's) is low and the corresponding doublet separation is large. The potential barrier preventing inversion in CH_3X (perhaps most easily visualized as an internal rotation of a CH_2 group relative to the rest of the molecule) would be expected, on the basis of ordinary chemical concepts, to be much higher than the barrier for inversion in NH_3, and the doublet separation would be expected to be much smaller. From (11), we see that the absorption becomes

$$\alpha_\nu=\beta'\nu^2\Delta\nu/(\nu^2+\Delta\nu^2), \qquad (17)$$

for $\nu\gg\nu_0$, which is independent of the exact value of ν_0. The experimental results can be fitted to an equation of this form if we assume $\Delta\nu$ proportional to the pressure. The selection rules, statistical weights, and matrix elements are the

same as in the ammonia problem. β' is then given by

$$\beta' = \frac{8\pi^2 N}{3kT}\frac{\mu_0^2}{Q}\sum_{J,K} g(J,K) \times \frac{K^2}{J(J+1)} \cdot e^{-W(J,K)/kT}, \quad (18)$$

if, in (17), the ν's and $\Delta\nu$'s are in cm^{-1}.

The experimental values of α_ν were corrected for the contribution arising from rotational absorption; this contribution was estimated from the theory to be 0.00004 for CH_3Cl and 0.00005 for CH_3Br at $\nu = 0.81$ cm^{-1} and negligible in other cases. The experimental values of β' and $\Delta\nu$ were computed from the corrected values of α_ν for $\nu = 0.81$ cm^{-1} and 0.32 cm^{-1} and $P = 1.0$ atmos. The variation of α_ν with P is reproduced satisfactorily by Eq. (17) with these values of the constants. The experimental values of $\Delta\nu$ are of the correct order of magnitude. The theoretical values of β' are 30–40 percent higher than the experimental values; this agreement may be regarded as sufficiently exact to verify the correctness of the assumed theory. The results are compared in Table III; it is to be noted that the variation of β' is not identical with the variation in μ_0 since the sums in (18) are appreciably different for the different molecules.

Since we have assumed the doublet separation to be negligibly small, and the frequency associated with the absorption to be essentially zero, we should arrive at the same result if we simply consider the molecules to be rigid rotators, characterized by the quantum numbers J and K. Now a non-rotating dipole molecule can absorb energy from an electromagnetic field because of the effect of collisions; the absorption is given by the Debye formula (3), which can be derived from classical theory. The mechanism can be visualized as the following: After each collision the dipole has a tendency to be oriented parallel to the field; between collisions the dipole is stationary. If the field is varying with time, the net effect of this mechanism is the production of a polarization which lags behind the field. The polarization current then has a component parallel to the field, resulting in an absorption of energy. The dipole moment vector of a symmetrical top molecule such as CH_3Cl, in the state characterized by J and K, can be resolved into a component parallel to the direction of the total angular momentum vector and a component perpendicular to this direction. The former component is stationary in space, and hence gives a full contribution to the Debye absorption; the latter component has a time average equal to zero and hence gives no Debye absorption. The square of the component of the permanent dipole moment μ_0 parallel to the total angular momentum vector is, since the dipole moment vector is along the symmetry axis, equal to $M_K^2\mu_0^2/M^2$ where M_K is the angular momentum about the symmetry axis of the molecule and M^2 is the total angular momentum. From quantum mechanics $M_K^2 = K^2h/2\pi$, $M^2 = J(J+1)h/2\pi$; the square of this component is thus equal to $K^2\mu_0^2/J(J+1)$. If, in Eq. (3), we replace μ_0^2 by this last expression, and sum over all values of J and K weighted by the proper statistical factors, we arrive exactly at Eqs. (17) and (18). In other words, the quantum-mechanical theory and the essentially classical theory of Debye lead to identical results. This is no accident, since the Debye theory is the limiting form of the general theory as the resonant frequency approaches zero.[2] It should be emphasized that the Debye theory cannot be applied in the form (3), the essential modification being that μ_0^2 in (3) must be replaced by the square of that component of μ_0 which is stationary in space. Although the two viewpoints concerning the CH_3X absorption are formally equivalent, the Debye picture is perhaps more in agreement with our conventional ideas.

TABLE III.

	$\Delta\nu$ (exp)	β' (exp)	β' (theor.)	$\mu_0 \times 10^{18}$
CH_3F	0.20	0.0053	0.0067	1.81
CH_3Cl	0.22	0.0037	0.0054	1.87
CH_3Br	0.21	0.0029	0.0038	1.78

COS

From the known moment of inertia of the linear molecule COS, namely, 1.37×10^{-38} g-cm^2, the wave-length for the transition $J = 1 \leftrightarrow J = 2$ is very nearly 1.24 cm, the operating wave-length. The wave-length is given by

$$\lambda_0 = \frac{4\pi^2 cI}{h(J+1)}. \quad (19)$$

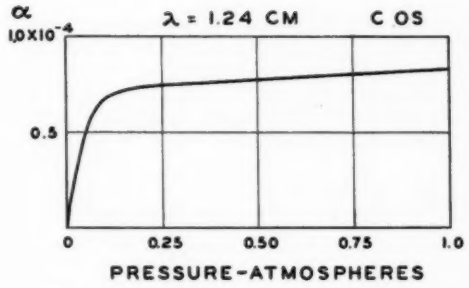


FIG. 5. Absorption of carbonyl sulphide.

The computed absorption is given by

$$\alpha_0=\frac{2\mu_0^2\nu_0^2h^2}{3cI(kT)^2}\times\frac{N}{\Delta\nu}, \tag{20}$$

where μ_0 for COS is 0.65 Debye units, ν_0 is 2.4×10^{10} cycles per second, and N is the number of molecules per cc. At one atmosphere, we obtain the observed value for α_0, namely 0.8×10^{-4} neper 1 cm, when $\Delta\nu$ is taken to be about 3×10^9 or 0.10 cm^{-1}. The form of Eq. (20) shows that if both N and $\Delta\nu$ vary linearly with pressure, α_0 is independent of pressure. The curve in Fig. 5, showing the absorption of COS, is of the resonant type as expected. In view of the above mentioned result $\Delta\nu=0.067$ cm^{-1} on the rather similar molecule N_2O, this seems to be a very reasonable value, and indicates a verification of the theory.

H_2S

A survey of the known rotational levels of H_2S[13] reveals no energy separations in the region $0<\nu<3.0$ cm^{-1} which correspond to an allowed transition. The microwave absorption in H_2S should, therefore, be very small, as is observed. The small absorption at $\nu=0.81$ cm^{-1} ($\alpha_\nu\times10^4\cong0.3$) is probably the cumulative effect of the numerous lines at higher frequencies.

SO_2

The rotational levels of SO_2 are not known experimentally, so a detailed calculation is not possible. The levels can be computed if the molecule is considered to be a rigid unsymmetrical top. It is found that there are, up to $J=12$, about 10 lines in the frequency range $0<\nu<1.50$ cm^{-1} which correspond to allowed transitions. At $J=12$ the Boltzmann factor is about 0.80, so there will undoubtedly be many strong lines in this frequency range arising from transitions involving levels higher than $J=12$. It is, therefore, reasonable to assume that the observed absorption in SO_2 could be accounted for on the basis of ordinary rotational transitions, indeed, no other mechanism seems possible.

[13] P. C. Cross, Phys. Rev. **47**, 7 (1935).

OTHER MOLECULES

The other molecules listed in Table I are non-planar unsymmetrical tops. As we proceed from a symmetrical top to an unsymmetrical top, we find that each K level splits into two levels, and a transition between these two levels is generally allowed. Transitions of this type, corresponding to $\Delta J=0$, $\Delta K=0$ in the symmetrical top, will produce a series of lines from $\nu\cong0$ upward. Also, certain transitions corresponding to $\Delta K\neq0$ are allowed in unsymmetrical tops. The moments of inertia of these molecules are large; therefore, transitions corresponding to $\Delta J=\pm1$ will involve low frequencies. The net result is that we have a very complicated rotational spectrum, containing many lines with frequencies from $\nu\sim0$ upward, and thickly scattered throughout the microwave region. Any exact calculations are almost impossible; however, we should expect the magnitude of the absorption to be about the same as for the methyl halides. The pressure curves often show a behavior intermediate between the resonant and non-resonant types, which is accounted for by the distribution of lines described above. Figure 6 shows such curves for ethylamine.

In general, one would expect all heavy non-planar unsymmetrical top molecules which have a large dipole moment to show an appreciable absorption in the microwave region. Since the selection rules are more stringent for planar and linear molecules, each molecule must be considered individually. Non-planar symmetrical top molecules should be analogous to the methyl halides.

DIELECTRIC CONSTANTS

Van Vleck and Weisskopf[2] have given the appropriate formula for the computation of that part of δ which arises from the permanent dipole moment μ_0. To this must be added the contribution arising from the electronic polarization P_A. For zero frequency, δ is given by

$$\delta_0=\frac{3(P_E+P_A)}{V}+\frac{4\pi N\mu_0^2}{3kT}, \tag{21}$$

where V is the molar volume.[14] δ should not differ very much from δ_0 in the microwave region, since this is on the low frequency side of the region of large absorption. This is illustrated by the comparison of calculated static values and observed values in Table II. The necessary data were taken from Smyth's text,[14] except for the substituted methanes, for which the data were taken from later papers by Smyth and his co-workers.[15]

For the methyl halides, the contribution to δ arising from the frequencies near zero (the ones to which we attributed the absorption) is given by

$$\delta_\nu' = \frac{4\pi N\mu_0^2}{3kT}\frac{\sum}{Q}\frac{\Delta\nu^2}{\nu^2+\Delta\nu^2}, \tag{22}$$

where $\sum$ is the sum in the numerator of (18). For the complex part of the dielectric constant we have

$$\epsilon_\nu'' = \frac{\alpha_\nu c}{2\pi\nu} = \frac{4\pi N\mu_0^2}{3kT}\frac{\sum}{Q}\frac{\nu\Delta\nu}{\nu^2+\Delta\nu^2}. \tag{23}$$

If we assume that the contribution to δ arising from the rotational frequencies is equal to the static value of this contribution, the above equations lead to

$$\delta_\nu = \delta_0 - \epsilon_\nu''\frac{\nu}{\Delta\nu} = \delta_0 - \frac{\alpha_\nu}{2\pi\Delta\nu}. \tag{24}$$

Insertion of the observed values of α_ν for CH_3Cl gives $\delta\times10^3=9.9$ at $\lambda=3.18$ and $\delta\times10^3=9.7$ at $\lambda=1.24$, as compared with the observed values 9.4 and 9.9. This agreement is not too good, but perhaps is within the limits of error of the measurements. The observed increase in δ as λ is decreased from 3.18 cm to 1.24 cm if real, must be attributed to the effect of the absorption lines at higher frequencies. As the frequency is increased from zero upward, the δ value of NH_3 should first increase above the static value and then decrease as the microwave absorption band is approached. The observed values are both lower than the static value; the meaning of this result is uncertain.

ACKNOWLEDGMENTS

We wish in particular to express our indebtedness to George W. Leck, Jr., for invaluable help in building microwave generating and power measuring equipment and to Elizabeth T. Bush for constructing wave guide components, preparing gases in the laboratory not readily available commercially, and taking many measurements. Also we express our appreciation to Irving Wolff of these Laboratories for support and encouragement, and to John Turkevich of Princeton University, for critical and helpful discussions during the course of the investigation.

[14] C. P. Smyth, *Dielectric Constant and Molecular Structure* (Chemical Catalog Company, New York, 1931).

[15] C. P. Smyth and R. H. Wiswall, Jr., J. Chem. Phys. 9, 356 (1941); C. P. Smyth and K. B. McAlpine, J. Chem. Phys. 1, 190 (1933); and J. Chem. Phys. 2, 499 (1934).

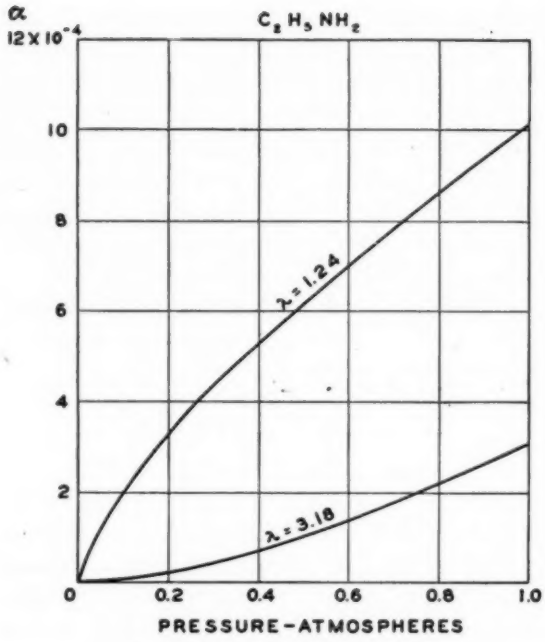


FIG. 6. Absorption of ethylamine.

Note added in proof, July 9, 1946.—Figure 4 in the paper referred to in reference 1 shows the absorption of ammonia at reduced pressure at three different wave-lengths. The information in this figure is quite consistent with the map of the fine structure of ammonia published by Bleaney and Penrose in the March 16, 1946, issue of *Nature*. The curve in Fig. 4 at 1.243 centimeters rises steeply at low pressure, but, after a few millimeters of pressure are reached, bends over and rises less steeply. A curve of this character indicates that the observations were taken near the maximum of one of the spectral lines. True enough, Bleaney and Penrose map an unresolved pair of lines at this wave-length to which they assign the quantum numbers (4,4) and (10,9). The curve at 1.320 centimeters also exhibits a point of inflection and these observations were taken on the complex and unresolved absorption line designated (5,4), (4,3), and (6,5). By contrast, the experimental curve at $\lambda=1.227$ is linear which is consistent with the fact that this wave-length falls *between* the (4,4) and (5,5) lines.

Electrical Observations of the Austenite-Martensite Transformation in Steel

ANDREW W. MCREYNOLDS
*Naval Proving Ground, Dahlgren, Virginia**

(Received March 14, 1946)

A method is described for observing the austenite-martensite transformation in steel by measuring electrical resistance. Small wire samples are heated and cooled either in vacuum or in air while temperature and resistance are automatically recorded. Temperature throughout the sample is uniform enough to permit a determination of resistance as a function of temperature. It is shown that from these data a complete record of transformation during cooling can be computed. Such transformation data have been obtained on three steels of widely different compositions, representing the types of steel in which transformation is most difficult to observe.

I. INTRODUCTION

AT temperatures above about 750°C, the stable phase of steel is the face-centered austenite. Relatively slow cooling causes a redistribution of carbon by diffusion and a readjustment of the lattice to the body-centered structure stable at lower temperatures. If cooling is rapid enough, however, the steel will instead transform without diffusion to a metastable tetragonal structure known as martensite, which is much harder than the products of transition by diffusion.

Study of the austenite-martensite transition is important in ferrous metallurgy, but its properties are such that observation is experimentally difficult. Some of the properties on which there is general agreement are:

1. Plate-like crystals of martensite form almost instantaneously, probably by a shearing process, with no diffusion.
2. The amount of transformation which has taken place at each temperature during cooling is a function of that temperature alone and is independent of the rate of cooling.
3. Transformation begins only when a definite temperature (referred to as the *Ms* point) has been reached, and this point cannot be changed by increasing the cooling rate.
4. The *Ms* point is depressed, approximately linearly with the amount added, by the addition of alloying elements.

Two general experimental methods can be applied to quantitative observation of the transformation: metallographic examination after heat treatment, or direct observation during transformation by simultaneously measuring temperature and some physical property connected with the structure. The latter method requires that temperature be uniform over at least a portion of the sample, and that some physical property of that portion be measured continuously during cooling and correlated with the structure.

Bain and Davenport[1] first used the metallographic method in 1930 to study isothermal transitions to pearlite or bainite, and to derive their well-known "*S*" curve. Transformations to martensite could not be observed by that method, but in 1940 Greninger and Troiano[2] slightly modified the technique of heat treatment by interrupting quenching of the sample at some temperature *T* with a brief reheating to temper the martensite formed above temperature *T*, and render it distinguishable from that formed below *T*. Their results indicated the properties of the transition listed above. The modified *S* curve based on those properties is shown in Fig. 1. Payson and Savage[3] applied the same "quench-temper" method to a study of the effect on the *Ms* point of alloying elements in steel.

Direct measurements have been made by a number of investigators using such phenomena as thermal arrest, dilatation, magnetic changes

* This article comprises part of a dissertation submitted to the University of California in partial fulfillment of the requirements for the degree of doctor of philosophy. Experimental work was carried out at the Naval Proving Ground, Dahlgren, Virginia.

[1] E. S. Davenport and E. C. Bain, Trans. A.I.M.E. **90**, 117 (1930).
[2] A. B. Greninger and A. R. Troiano, Trans. A. S. Metals **29**, 537–574 (1940).
[3] P. Payson and C. H. Savage, Trans. A. S. Metals **33**, 261–275 (1943).

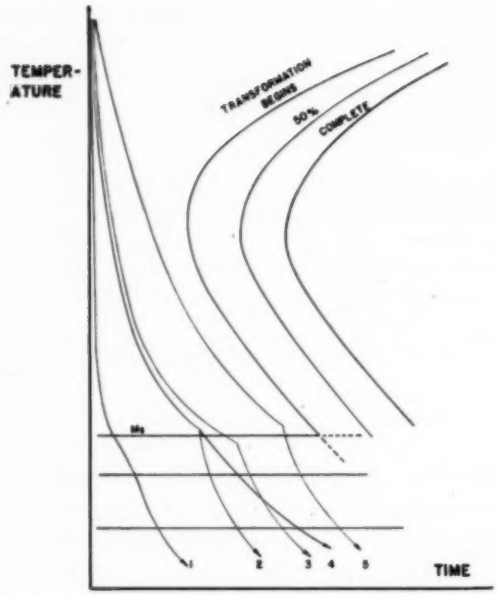


FIG. 1. Typical transformation S curves for steel and cooling curves for test specimens. (1) Rapid gas quench, (2–4) initial cooling in vacuum, gas cooling through transition region, (5) slow initial cooling.

and the noise of structural change. These measurements have been limited, however, to determination of the *Ms* point only, or to measurements at and below room temperature, on steel of such composition that it remained mostly austenite until cooled below room temperature.

Temperature-time measurements on small samples quenched under various conditions were made by French and Klopsch,[4] Esser, Eilender, Riedel, and Spenle,[5] Wever and Engel,[6] Wever and Rose,[7] and by Greninger.[8] Temperature was measured by galvanometers with periods of 0.020 second or less. The *Ms* point as determined by the thermal arrest was found independent of the cooling rate in all cases.

Tamann and Scheil[9] and later Scheil[10] made dilatometric measurements in the temperature region below 20°C on high carbon steel. They found transformation to depend only on temperature. Förster and Scheil[11] made both acoustical and electrical resistance measurements in the same temperature region and found the time of formation of a single martensite plate to be of the order of 0.02 second.

All of the above methods of direct observation were limited to rather special conditions and were not adequate to determine complete data on the progress of transformation. The present paper describes a method for recording electrical resistance and temperature and for interpreting those data to furnish a complete record of the amount of transformation at all stages of the cooling. The technique of heating and cooling was such that at any instant the temperature throughout the sample was uniform within a few degrees. Electrical resistance is a suitable physical property to provide a measure of the proportions of the constituents present because it can be easily and rapidly measured and recorded at all temperatures, and because, as will be shown below, it can be quantitatively correlated with structure. The advantages of the electrical method over metallographic methods are:

1. It yields precisely measurable data and requires no visual estimates of relative proportions of constituents.
2. Measurements can be made directly during a normal cooling operation.
3. It is applicable to steels of almost any composition.
4. A complete record of percent of the sample transformed *vs.* temperature can be determined from the cooling of one specimen.

II. APPARATUS

Two different types of apparatus were used, one for heating and cooling the samples in vacuum and one for heating and cooling in fluid media. The first method has the advantages of more precise control of temperature and freedom from surface contamination or decarburization of the sample. The second method has the advantages of greater ease in preparing and mounting samples and a wider range of applicability to different compositions of steel. Since

[4] H. J. French and O. Z. Klopsch, Trans. A. S. Steel Tr. **6**, 251–293 (1924).
[5] H. Esser, W. Eilender, K. Riedel, and E. Spenle, Archiv. f. Eisenh. **4**, 113–144 (1930).
[6] F. Wever and N. Engel, Archiv. f. Eisenh. **5**, 377–382 (1932).
[7] F. Wever and A. Rose, Mitt. Kaiser Wilh. Inst. f. Eisenf. **19**, 289–298 (1937).
[8] A. B. Greninger, Trans. A. S. Metals **30**, 1–26 (1942).
[9] G. Tammann and E. Scheil, Zeits. f. anorg. Chemie **157**, 1–21 (1926).
[10] E. Scheil, Zeits. f. anorg. Chemie **183**, 98–120 (1929).
[11] Förster and E. Scheil, Zeits. f. Metallkunde **28**, 245–247 (1936).

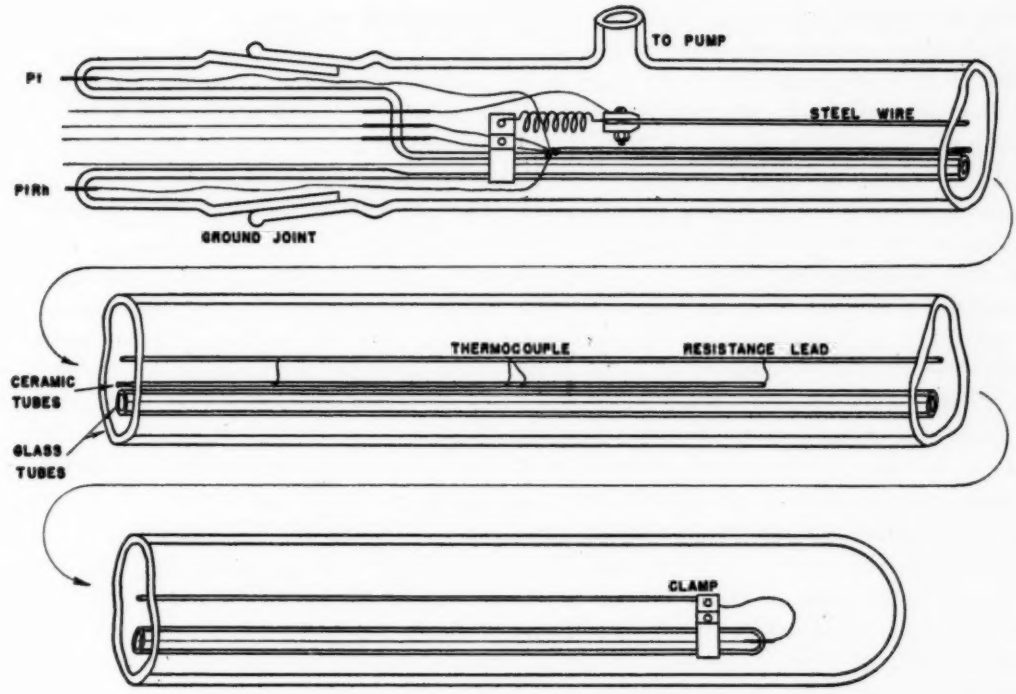


FIG. 2. Vacuum chamber, diameter 1.5 in., and steel wire sample.

cooling can be accomplished much more rapidly in fluid media than in vacuum, the second method is applicable to steels which require rapid cooling to cause transformation to martensite.

1. Apparatus for Heat Treatment in Vacuum

The samples were cut from heavy stock and drawn into uniform circular wire of 0.026-inch diameter. This size was chosen as small enough to allow cooling at rates up to 150°C per second without excessive radial temperature gradients and at the same time large enough to allow the attachment of 0.002-inch thermocouple leads without greatly altering the metal temperature near the junction. Segments of wire 50 cm long were polished with fine emery cloth, cleaned with carbon tetrachloride, and suspended under slight spring tension between clamps mounted on a 7-mm glass tube (see Fig. 2). The complete sample holder was inserted in a larger glass tube evacuated to 10^{-5} mm Hg. Heating was accomplished by passing 60-cycle alternating current directly through the sample. Cooling at rates of from 1°C per second to 200°C per second could be accomplished either by radiation alone, or by transfer of heat through a small amount of helium, of the order of 1-mm Hg pressure, admitted for that purpose. The wire surface was carefully polished since the heat emission constant of the surface had to be uniform to preserve axial uniformity of temperature.

Temperature measurements were made by the use of a thermocouple of two 0.002-inch wires, platinum and platinum-10 percent rhodium, spot-welded to opposite sides of the sample at its center. (Fig. 3.) The thermocouple leads could be connected by a selector switch to either a Rubicon Precision Type B potentiometer or a galvanometer. The galvanometer was a Leeds & Northrup Type R, with 500-ohms coil resistance, which had been modified by the installation of a 0.002-inch round platinum-10 percent rhodium suspension to have a period of about 1 second. Its deflections, recorded photographically on a drum of sensitized paper moving at 1 cm per second, were found to be reproducible within 0.1 mm, corresponding to about 1°C. (A rapid recording potentiometer of the "Speedomax" type would be satisfactory for temperature recording.)

Resistance was measured between two 0.002 inch platinum-10 percent rhodium wires, one spot welded to the sample 7.5 cm on each side of the thermocouple leads. Measurements were thus taken only on the central 15 cm of the wire, and 17.5 cm were allowed on each end to eliminate non-uniformity of temperature due to conduction to the supports. Temperature uniformity was verified before each run by measuring the thermoelectric voltage between each of the resistance leads and the central platinum thermocouple wire and comparing with the voltage of the central thermocouple. To measure resistance, a constant direct current of 0.5 amp was passed through the wire and a galvanometer of the same type as used for temperature measurements was connected across the 0.002-inch resistance leads. Resistance and temperature measuring galvanometers recorded on the same moving film. A calibration was recorded on each run by connecting the resistance-measuring galvanometer across a standard resistance in series with the sample, and timing lines were marked on the paper by a lamp flashing at 1-second intervals.

Procedure in making a run was: The sample was heated in vacuum to 900°C and held at that temperature for 10 minutes to permit all carbon to go into solid solution. (A longer time was found to have no effect on the *Ms* temperature.) The heating current was then cut off to start the cooling. The time required in vacuum to reach the transition temperature (260°) was about 140 seconds. This time could be reduced to as little as 4 seconds by the admission of a little helium to the system, or could be made as long as desired by reducing the heating current gradually. A small quantity of helium was always admitted to the system when the temperature was a few degrees above or a few degrees below the *Ms* point, in order to accelerate cooling through the transition region since cooling by radiation alone would have been extremely slow at low temperatures. The time required to cool through the transition temperature range was thus made too short to allow either the formation of diffusion products (e.g., bainite) from the austenite or an appreciable tempering of the newly-formed martensite. After the sample had cooled to room temperature, the chamber was again evacuated and the sample reheated to 200°–250°C in a time of 30 seconds. Since resistance could not be measured with the heating current on, heating was carried out in steps of about 20°, with the heating current turned on for one second and then off for three seconds while the measurements were made.

Both temperature and resistance were continuously recorded during the cooling and reheating operations. A typical record is shown in Fig. 5. Each sample was used for about ten records under various cooling conditions and was then divided into sections for carbon analysis, microscopic examination, and hardness tests.

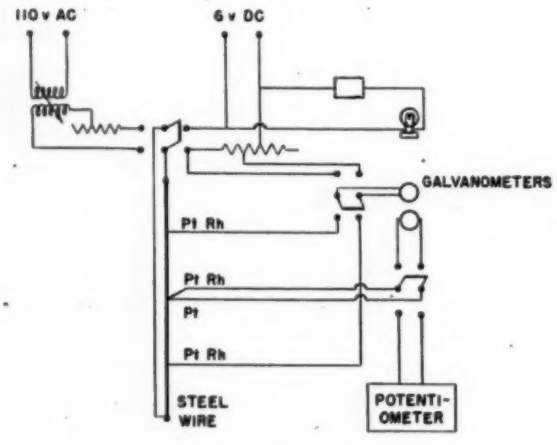


FIG. 3. Electrical circuit diagram.

2. Apparatus for Heat Treatment in Fluid Media

Bars 3 to 4 inches long and varying in cross-sectional size and shape from 0.025-inch diameter circles to 0.060-inch squares were cut from the sample material. Each bar was bent into a U-shape and supported by 36-gauge Chromel and Alumel wires spot welded near the ends of the U as shown in Fig. 4. The one Chromel and five Alumel wires served both as supports and as current, potential, and thermocouple leads. Each wire ran through a separate hole in a bundle of ceramic tubes and connected to one element of a flexible 6-conductor cable leading to the recording apparatus. The same recording instruments as described in the preceding section measured the thermoelectric voltage and the potential drop across a 1-inch segment of the sample caused by direct current of 0.5 ampere

which was passed through the sample. Thermocouple leads were attached outside the segment through which current passed, in order to avoid spurious potential drop due to the measuring current.

Heating was accomplished by inserting the sample (but not the ceramic tube support) either into a small lead bath 1 inch in diameter by 8 inches deep, heated to 900°C in an electric furnace, or into a hollow ceramic tube $\frac{3}{4}$ inch in inside diameter by 10 inches high, which was closed at the bottom and heated in the same furnace as the lead bath. After remaining at 900°C long enough for carbon to go into solid solution, the sample was cooled by withdrawing it into still air, or if rapid cooling was required, into a blast of air from a high speed fan mounted just above the heating furnace. After initial cooling by air blast, the sample was removed to still air to cool more slowly through the transition region. This cooling procedure was found to be rapid enough to quench a 0.025-inch round sample of plain carbon (0.88 percent) steel wire from 900°C without allowing any observable transformation at temperatures above the *Ms* point of 200°C. Although in that case time to cool from 900°C to 300°C was only 2.5 seconds, satisfactory temperature and resistance records were obtained during slower cooling below that temperature. Cooling rates at the beginning point of transformation were of the order of 20°C per second for the smallest samples used.

An alternate cooling method tried was to withdraw the sample from the heating furnace, immediately plunge it into a small bath of molten $NaNO_2$ and KNO_3 at 300°C, then allow it to cool with the salt bath while continuously stirring. This procedure was found to have no advantages over air cooling and to be much less convenient.

Some samples were cooled in air down to 120°C and then immersed in boiling water (at 100°C to restore complete uniformity of temperature. They were next immersed alternately in water at room temperature (27°C) and at 100°C several times, being left in each bath long enough for a resistance measurement to be taken. Data thus obtained provided an accurate measurement of the amount of transformation between 100°C and 27°C, and of the temperature-resistance relation of the final structure. The method of interpretation of data is explained in the following section.

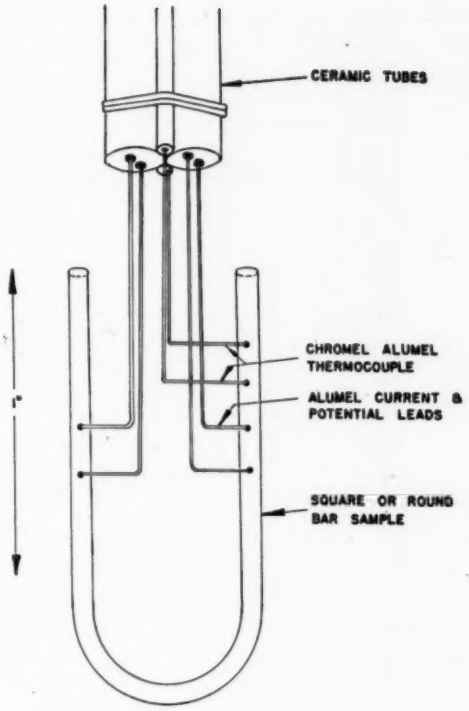


FIG. 4. Sample mounting for heating in furnaces and liquid baths and cooling in air.

III. INTERPRETATION OF DATA

A sample photographic record, obtained on cooling in vacuum and helium, is shown in Fig. 5. Galvanometer deflections were read to an accuracy of 0.1 mm, corresponding to 0.1–0.3 percent of the resistance and to about 1°C in temperature. Thermocouple deflections were converted to degrees C, but resistance measurements were left in arbitrary units. The data thus obtained could be interpreted to determine the amount of transformation at any stage of the cooling and also the heat of transformation.

The data were plotted (see Figs. 6–8) as resistance-temperature curves which exhibit a sharp break in slope at the *Ms* point. An essential condition of the procedure was that temperature must be uniform throughout the sample during cooling, since even a slight non-uniformity could give a spurious shape to the resistance-tempera-

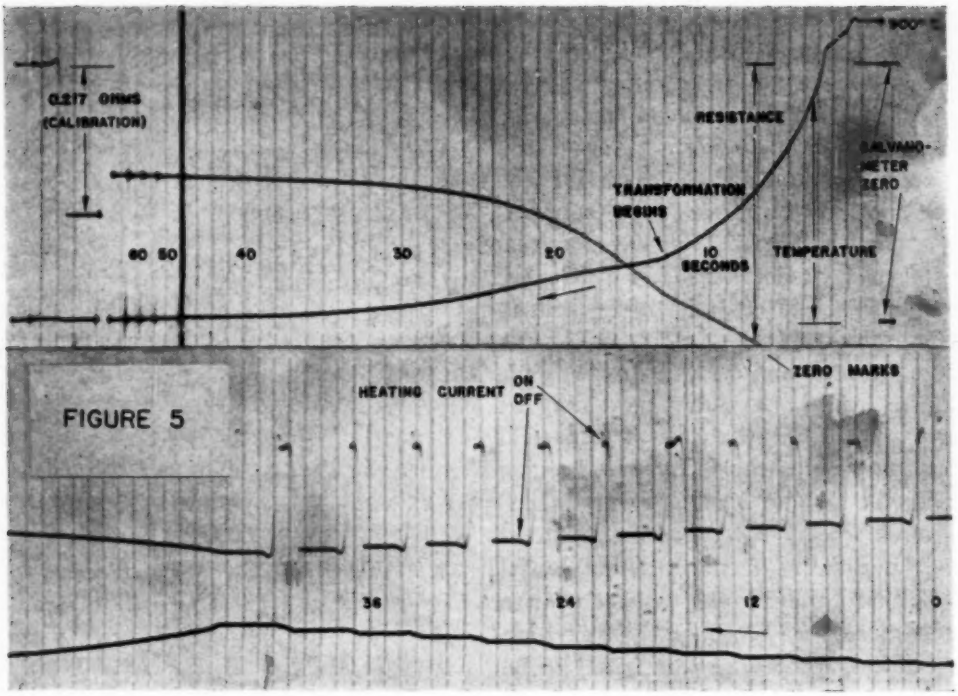


FIG. 5. Typical record of temperature and resistance. Top portion represents cooling and bottom portion reheating.

ture curve. The sharpness of the break at the *Ms* point could, however, be used as a criterion of uniformity, and all records in which the break was blurred into a round corner were discarded.

The curves in Figs. 6–8 will be seen to consist of two linear portions and a connecting curve. The two linear segments represent the resistance-temperature relation when the crystal structure is 100 percent austenite and 100 percent martensite, respectively, and the connecting curve represents the resistance as a function of temperature of a mixture of the two constituents as their proportions continuously change with decreasing temperature. If it is assumed that the linear austenite resistance curve can be extrapolated below the *Ms* point without much error to show the resistance of austenite retained in that temperature region, then from the position of the connecting curve between the linear segments, the proportions of the constituents present at any time and temperature can be determined. The resistance of the mixture depends on both the amount and the distribution of the martensite plates in the austenite matrix. The extreme possible distributions are with the martensite either entirely in series with the austenite or entirely in parallel. These distributions lead, respectively, to the relations:

$$\frac{\%\ \text{Martensite}}{100}=(R_a-R)/(R_a-R_m)\ \text{(series)},$$

$$\frac{\%\ \text{Martensite}}{100}=\frac{R_a-R}{R_a-R_m}\times\frac{R_m}{R}\ \text{(parallel)},$$

where

R = resistance of the mixture.
R_a = resistance of 100 percent austenite.
R_m = resistance of 100 percent martensite.

For a ratio $R_a/R_m=1.5$, these relations differ by a maximum of 10 percent as shown in Fig. 9. Actually, the martensite is known to be distributed as randomly oriented small plates throughout the matrix. It was considered a sufficiently close approximation to assume that the martensite forms in thin parallel plates at 45 degrees to the direction of current flow and extending completely across the sample. It can easily be shown that with such a distribution $R=\frac{1}{2}(R_{\text{series}}+R_{\text{parallel}})$. On the basis of this

assumption, values of the function $(R_a-R)/(R_a-R_m)$ were computed and are shown in Fig. 9. The amount of transformation was determined from the resistance-temperature curves by first finding the ratio $X=R_a/R_m$ then using the appropriate curve from Fig. 9 to find the per-cent martensite corresponding to the value of $(R_a-R)/(R_a-R_m)$ at any desired temperature. A number of resistance-temperature-time curves were obtained on each sample, and for each curve Fig. 9 was applied to determine the temperatures at which transformation was 10, 25, 50, 75, and 90 percent complete.

Since the *Ms* point was independent of the rate of cooling and since microscopic examination showed the final structure to be martensite it is clear that the deviation of the curve from linearity represented a transformation, but it still had to be determined whether the shape of the curve was influenced by other reactions. The fact that no such reactions occurred above the *Ms* point is evidenced by the linearity of the resistance-temperature relation in that region. To investigate the possibility of tempering of the newly-formed martensite below the *Ms* point, a number of samples were cooled in

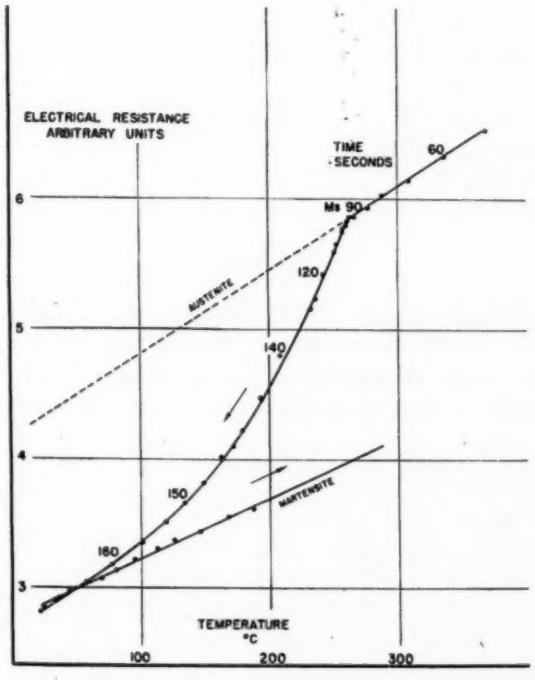


FIG. 6. Resistance-temperature curve for Ni-Cr alloy steel cooled in vacuum for 130 seconds, then more rapidly in helium. Position of non-linear segment indicates degree of transformation from austenite to martensite.

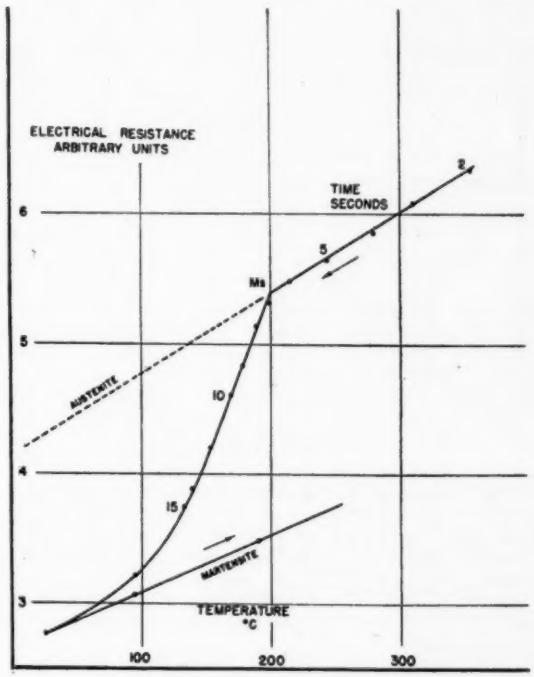


FIG. 7. Curve for 4.5 percent Mn low carbon steel cooled rapidly in air blast.

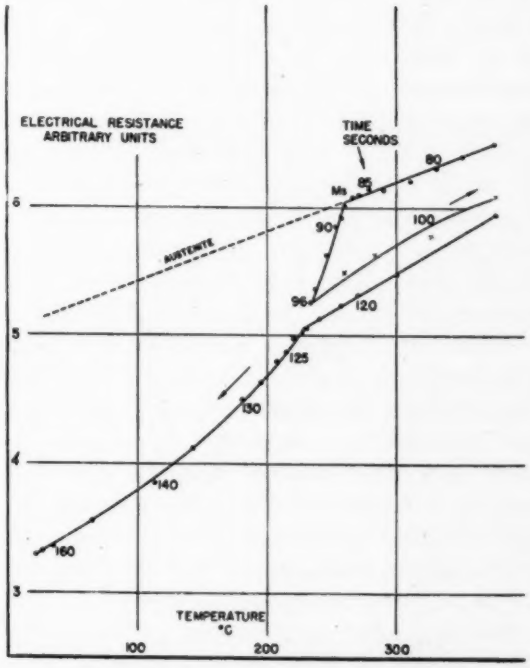


FIG. 8. Curve for Ni-Cr alloy steel. Cooling was interrupted at 130°C. Transformation stopped at that point and resumed when temperature next dropped below 130°.

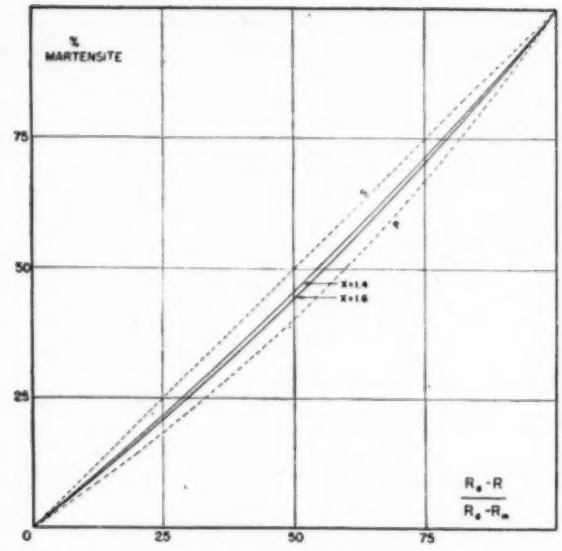


FIG. 9. Variation of resistance of an austenite-martensite mixture with proportions of the constituents. Dotted curves show variation with the constituents in series or in parallel with each other, solid curves with constituents assumed to be in alternate thin plates at 45° to the direction of current, $X = R_a/R_m$, R = resistance of mixture, R_a = resistance of 100 percent austenite, R_m = resistance of 100 percent martensite.

vacuum nearly to the *Ms* point, then at various rates through the transition region. Figure 10 shows that the computed amount of transformation as a function of temperature is the same for all of these tests. It is therefore evident that time-dependent tempering reactions either did not occur or did not influence the determination of the amount of transformation. Local stresses set up by the formation of martensite plates might be expected to increase the resistivity of both the martensite and the surrounding austenite and thus influence the shape of the resistance-temperature curve. The amount of the increase in resistivity, however, should be not greater than would be caused by extensive cold work; that is around one percent for steel. Furthermore the increases in resistivity of austenite and martensite enter into the determination of the amount of transformation in such a way as to approximately cancel each other. It is therefore safe to assume that stresses accompanying transformation introduce no serious error.

Metallographic examination of the samples after cooling to 20°C showed them to be all martensite, as nearly as could be determined. The resistance curves also indicate that transformation had practically ceased at that temperature. It has therefore been assumed in the computations that the structure after cooling to room temperature was 100 percent martensite.

From the data of Table II and from a determination of the decrease of rate of cooling as transformation begins, a value of the heat of formation of martensite can be calculated. The rate of heat radiation is substantially the same immediately before transformation begins as immediately after, but after reaction begins part of this heat is supplied by the exothermic reaction so that the cooling rate is reduced to a value θ_2 determined by the equation:

$$\begin{aligned} ms\theta_1 &= ms\theta_2 + mHM\theta_2, \\ HM\theta_2 &= s(\theta_1 - \theta_2), \\ H &= \frac{s}{M}\left[\frac{\theta_1}{\theta_2} - 1\right], \end{aligned} \tag{1}$$

where

H = heat of reaction,
s = heat capacity of austenite,
m = mass under consideration,
M = fraction of martensite formed per degree decrease in temperature,
θ_1 = cooling rate just before transformation begins,
θ_2 = cooling rate just after transformation begins.

To determine H, only the ratio of cooling rates need be determined and therefore only the slopes of the temperature-time galvanometer record need be measured, and no calibration is required.

IV. RESULTS

Samples of the compositions listed in Table I were tested. Data on the progress of transformation with decreasing temperature are given in Table II and summarized in Table III. These

TABLE I

Sample Shape	0 0.026" wire	1 0.026" wire	2 0.060" wire	3 0.060" bar	4 0.025" wire
Element			Percent		
Carbon	0.41	0.41	0.38	0.09	0.88
Manganese	0.26	0.26	0.26	4.53	0.33
Molybdenum	—	—	—	0.93	—
Nickel	3.85	3.85	3.85	—	—
Chromium	1.85	1.85	1.85	—	—
Silicon	0.074	0.074	0.074	0.12	0.12
Phosphorus	0.015	0.015	0.015	0.014	—

results confirm a number of properties of the transformation which have been previously suggested on the basis of metallographic and other investigations.

The *Ms* point and also the amount of transformation which had taken place at a given temperature were found to be independent of both the over-all cooling rate and the rate of cooling through the transition region within the range investigated. Figures 10 and 11 show this property to hold for rates from 1°C per second to 100°C per second above the *Ms* point and for times between 10 and 75 seconds to cool from the *Ms* point to the point of 90 percent transformation.

Tests in which cooling was interrupted before the completion of transition showed that transformation stopped as soon as cooling was interrupted but resumed when cooling next proceeded below the point of interruption. Figure 8 is the record of one such test in which the sample was cooled to about 230°C, rapidly reheated to about 360°C, then allowed to cool. It will be seen that the resistance curve does not return to its original position at 230°C, but the discrepancy is attributable to tempering of the martensite during reheating.

From the chemical compositions of the three samples values of the *Ms* points were calculated according to the approximate formulas of Payson and Savage,[3] which were based on extrapolated metallographic data. The agreement between these predicted values and the experimental data, both shown in Table III, confirm that the

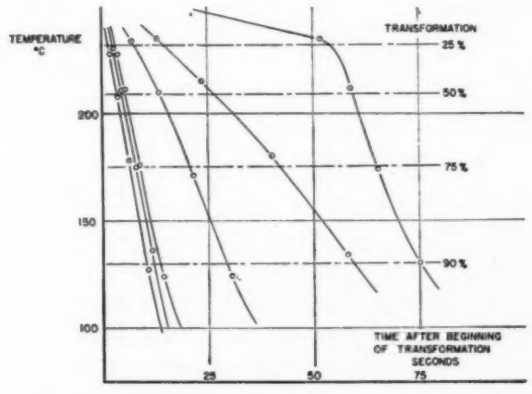


FIG. 10. Effect of cooling rate through the transformation region on the progress of transformation. Cooling curves are for Ni-Cr alloy steels.

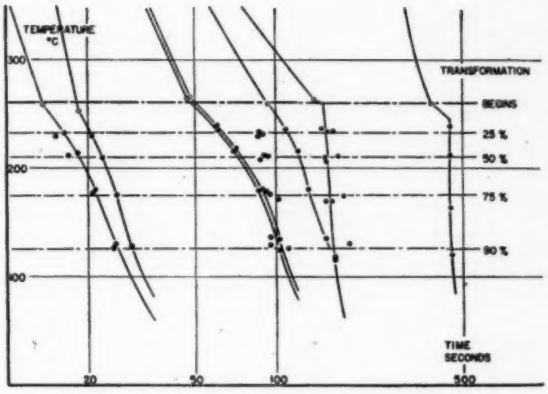


FIG. 11. Effect of cooling rate on the progress of transformation. Cooling curves are for a number of Ni-Cr alloy steel samples. Points represent individual determinations from the resistance-temperature curves of the temperature at which transformation reached various stages of completion.

Ms point is determined almost entirely by the chemical composition.

Although no tests were conducted for the specific purpose of measuring heat of formation of martensite from austenite, values of this quantity were determined from some of the records by using Eq. (1). Values of $s=0.115$ cal. per gram per degree centigrade and $M=0.0108$ and $M=0.0089$ per degree centigrade for samples 0 and 1, respectively, were used in the calculations.

V. CONCLUSIONS

1. From measurements of electrical resistance and temperature, continuously recorded by the method described, complete quantitative data on the transformation of a steel sample from austenite to martensite can be determined. The electrical resistance method is applicable to almost any type of steel, including the two types experimentally most difficult to study:

(a) Steel of low carbon content (less than 0.4 percent) which cannot be studied metallographically or by any other method yet developed.

(b) Rapid-reacting steels which must be quenched very rapidly to produce the martensite structure.

Satisfactory results have been obtained on both a 0.09 percent carbon steel and a 0.88 percent carbon plain carbon steel.

TABLE II. Transformation temperatures.

Sample	Record	Cooling medium: Initial temp. to *Ms* point	Cooling medium: *Ms* point to room temp.	Transformation: Begins (*Ms*)	10%	25%	50% °C	75%	90%
0	1	vacuum	helium	257	247	234	210	171	124
0	2	vacuum	helium, rapid	—	—	228	208	178	127
0	3	vacuum	helium, slow	258	248	235	215	180	134
0	4	vacuum	helium, rapid	—	—	231	211	175	136
0	5	helium, rapid	helium, rapid	—	—	228	211	176	124
0	6	vacuum	helium, slow	261	249	235	212	174	130
0	7	vacuum, slow	helium	258	250	238	212	163	118
1	1	vacuum	helium	262	247	229	194	147	100
1	2	vacuum	helium	264	249	227	200	150	95
1	3	helium, slow	helium	263	249	232	203	158	102
1	4	helium, slow	helium	263	245	229	200	152	95
1	5	vacuum	helium	264	253	237	209	168	114
1	6	vacuum	helium	264	251	234	207	168	114
1	7	helium, rapid	helium, rapid	250	241	229	208	173	126
1	8	helium, rapid	helium, rapid	256	246	232	211	178	128
1	9	helium, slow	helium	264	252	239	217	179	133
1	10	helium, slow	helium	262	249	234	215	178	123
Average				260	248	232	209	169	119
2	1	lead-salt, rapid	salt-slow	277	267	255	232	200	165
3	1	air, rapid	air, rapid	342	333	317	281	229	174
3	2	air	air	336	329	316	280	216	154
3	3	air, slow	air, slow	328	319	304	272	207	138
3 Average				335	327	312	278	217	155
4	1	air blast	air	200	189	173	150	123	93
4	2	air blast	air	192	181	167	147	121	99
4 Average				196	185	170	148	122	96

TABLE III.

Sample		0, 1	3	4
Composition		0.41C 3.85Ni 1.85Cr	0.09C 4.53Mn	0.88C 0.33Mn
Transformation:	Begins	260°C	335°C	196°C
	25%	232	312	185
	50%	209	278	170
	75%	169	217	148
	90%	119	155	96
Ms point calculated from composition		242°C	315°C	210°C

TABLE IV.

Sample	Cooling rate above *Ms* °C per second	Cooling rate below *Ms* °C per second	Heat of transformation cal. per g
0	21.0	7.1	25.2
0	12.8	5.25	18.6
0	21.5	9.3	16.9
0	22.1	8.7	20.8
1	26.2	11.05	17.6
1	29.3	11.5	20.0
Average			19.8
3	9.7	3.17	21.8
3	13.5	3.26	33.1
3	14.9	5.5	21.1
3	11.05	3.66	21.4
3	10.85	3.42	22.8
3	17.7	4.32	32.8
3	15.8	4.15	29.5
Average			24.5

2. Data on the alloy steel samples show that the temperature at which transformation begins and the amount which has been transformed at any temperature during cooling are independent of the time required for the steel to cool to that temperature (if no diffusion reactions occur), and hence, if cooling is interrupted before transformation is complete, the formation of martensite ceases and resumes only when the temperature is reduced below the point of interruption. These properties of transformation are at present generally accepted on the basis of previous direct measurements (during slow cooling below room temperature) and of metallographic observations.

The author is indebted to Mr. H. Elmendorf of the American Steel and Wire Company for drawing the samples into wire form, and to

Professor L. B. Loeb of the University of California, and Commanders B. R. Queneau, R. A. Sawyer, and R. N. Varney, of the Naval Proving Ground for helpful discussions. The opinions and assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

Energy Build-Up in Magnetrons

LLOYD P. HUNTER
Westinghouse Electric Corporation, Pittsburgh, Pennsylvania
(Received April 3, 1946)

An analysis of the operation of resonant cavity magnetrons is made without regard to the mechanism of conversion of the d.c. input power into the r-f output power. The magnetron is represented by its equivalent circuit and is assumed to be essentially in equilibrium at every point of its energy build-up curve. The last assumption is shown to be reasonable on the basis of measured starting times. From the data of load impedance charts and the rise of the r-f pulse, the relation between power generated and vane voltage is deduced. By considering the division of the energy between that stored in the resonant system, and that dissipated in the load, the law of build-up is derived. The dependence of starting time on load is calculated and agrees with experiment when the extra energy stored in the connecting line is taken into account. The starting time is affected slightly by the initial noise level and becomes infinite below a minimum Q. For high Q values the starting time can be varied only by changing the energy stored in the line, which may enable one to design a line for "mode" suppression.

INTRODUCTION

ONE of the most important parameters of magnetron performance is the radiofrequency voltage on the slots of the magnetron. This parameter is not directly accessible to measurement and has therefore been little considered in magnetron design. The purpose of this paper is to deduce an empirical law relating the generated power of a magnetron to the slot voltage, and then to show how this law may be used to calculate various features of magnetron performance which are of practical interest.

This investigation is specifically applied to resonant cavity magnetrons but illustrates a method of performance analysis which should be quite valuable when applied to any of a large class of high frequency generators. The purpose of the method is not to give any theory of the conversion mechanism in which the d.c. applied energy is converted into radiofrequency energy, but rather to give a performance analysis which should contribute greatly to the understanding necessary for intelligent engineering design of oscillators of the magnetron type.

First the slot voltage will be calculated for various values of load with a constant value of d.c. voltage applied to the magnetron. This will enable us to plot the output power delivered for these various loads against the slot voltage. Next we will supplement this curve of power output against slot voltage with similar data obtained from curves of the build-up of the r-f energy at the beginning of an applied d.c. voltage pulse. The curve will now extend over a wide range of slot voltages and we will proceed to fit an empirical relation to it.

The empirical relation so deduced will be used first to integrate the differential equation of energy build-up. The resulting equation can then be solved explicitly for the time, and we will proceed to calculate the variation of starting time as a function of load and of noise level. The form of the r-f build-up curve will also be calculated along with the conversion efficiency of the magnetron.

SECTION I

Relation of Power Generated to Slot Voltage

In this section the one fundamental assumption involved is that steady-state data, such as

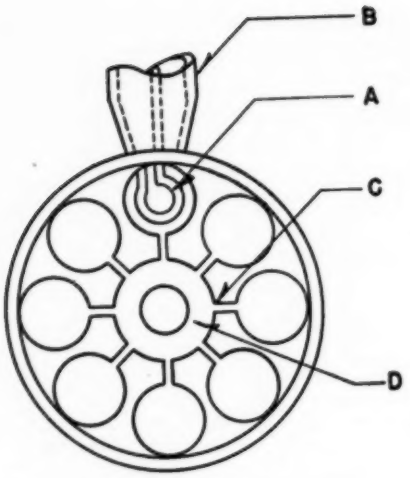


FIG. 1. Resonant cavity magnetron.

the variation of output power with load, may be applied to the build-up process. The results of starting time[1] measurements support such an assumption since they show that the complete build-up process occurs in about 300 cycles of the r-f. This means that the voltage is essentially constant for a few cycles and we are justified in postulating a state of quasi-equilibrium for our calculations.

In Fig. 1 is shown a diagram of a resonant cavity magnetron of the so-called "hole and slot" type. At A is shown the coupling loop which bleeds off r-f power and sends it down the coaxial line B to the load. At C is shown one of the slots across which the r-f voltage is developed by the circulating electron cloud in the interaction space D. Such a magnetron is typical, and although it may have a wave guide output rather than a coaxial line-coupling loop arrangement, it will serve for our discussion in either case.

One well-known method[2] of representing the performance of a magnetron is to plot contours of constant power out in the complex load impedance plane while holding the d.c. anode voltage constant. Such a load impedance chart is shown in Fig. 2. The solid contours are constant power contours while the dashed contours are constant operating frequency contours. From such data it is possible to calculate the manner in which the power output and power generated vary with the slot voltage. However, only the variation at high slot voltages is available in these data and one must look elsewhere for this variation near zero slot voltage. It is the purpose of this paper to show that such information can be obtained from measurements[1] of the build-up of the r-f envelope at the beginning of a pulse. To do this we will first examine the available data concerning the variation of output power with load, and then we will show how the measurements of r-f build-up allow us to extend these data.

First calculating the variation of output power with r-f voltage on the slots of the magnetron, we will assume a simple series equivalent circuit for the magnetron as seen from the back of the magnetron cavity. Now by definition

$$\text{Transmitted } Q = Q_T = \frac{2\pi \text{ energy stored}}{\text{energy out/cycle}} = \frac{2\pi V^2/\omega z_0}{P_0/f}, \quad (1)$$

giving

$$V = (P_0 Q_T z_0)^{\frac{1}{2}} \quad \text{(r.m.s. volts)},$$

where P_0 = output power and $z_0 = (L/C)^{\frac{1}{2}}$ for the magnetron equivalent circuit.

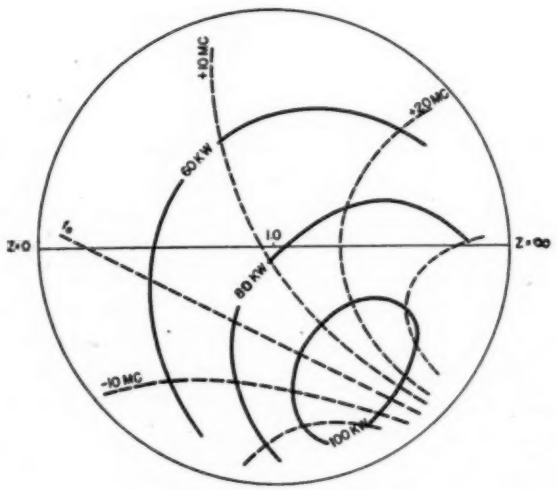


FIG. 2. Load impedance chart.

This formula gives us the relation between the r-f voltage on the slots of the magnetron and the power delivered to the load. We will not concern ourselves at this point with the power lost in the circuit, since such losses do not enter into the

[1] More and Miley, Massachusetts Institute of Technology Report 509.
[2] Evans and Rieke, Massachusetts Institute of Technology Report 52–2.

calculation of the slot voltage using Eq. (1), nor will they enter into the calculation of the build-up of stored energy with time, since that may be approached by using the output power as well as the power generated. We will consider all losses, however, when we come to calculate the total power generated as a function of slot voltage.

Referring the load impedance chart of a 2J32 tube[3] to constant d.c. voltage by means of a performance chart, we may calculate the curve of Fig. 3 by means of formula (1). Values of P_0 and corresponding Q_T values may be obtained directly from the load impedance chart if it is remembered that the Q varies directly with the voltage standing wave ratio, and that the region of low Q or heavy loading corresponds to the region of crowded frequency contours. Figure 3 is a plot of output power as a function of the square of the r-f voltage on the slots of the magnetron. As can be seen from Fig. 3, these data are not sufficient to tell us much about the low voltage end of this characteristic curve. However, the results[1] of More and Miley on starting time and build-up of the r-f pulse in magnetrons is sufficient to indicate how the curve of Fig. 3 approaches zero.

In Fig. 4 is shown a typical family of r-f

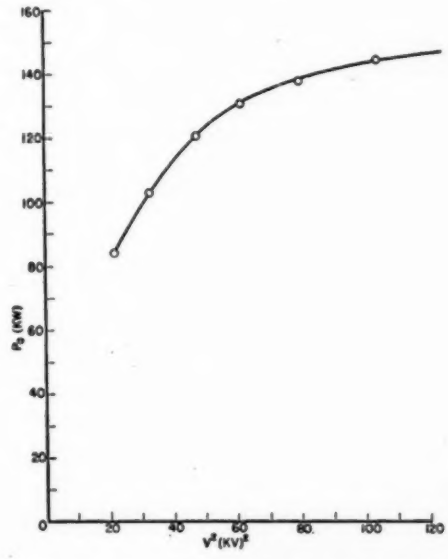


FIG. 3. Characteristic curve from load impedance chart.

[3] Platzman, Evans, and Rieke, Massachusetts Institute of Technology Report 451.

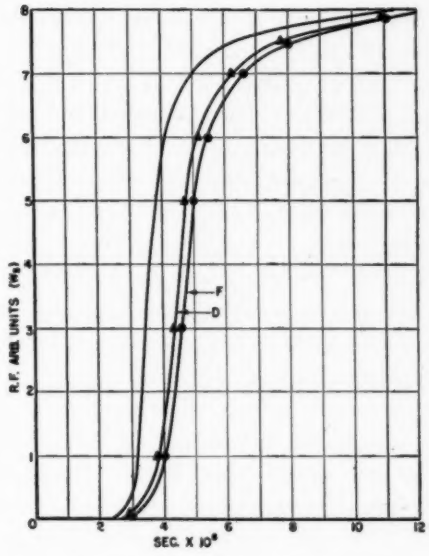


FIG. 4. Build-up curves from r-f envelope.

build-up curves. The curve marked d is plotted directly from the photograph d of Fig. 3 of reference 1. Similarly, curve f is plotted from photograph f. These two curves are for different rates of rise of the voltage pulse given in reference 1. The solid curve of Fig. 4 represents the response of the magnetron to a perfectly square voltage pulse of zero time of rise. This curve was obtained from the other two by a method devised by Dr. William Altar,[4] which gives the response of a general circuit to an ideal pulse of infinite rate of voltage rise, from the known response of the network to a pulse with a finite and constant rate of voltage rise. Since this method is an application of the inversion of Carson's theorem on a transient response, it presupposes proportionality of the r-f voltage response of a magnetron *vs.* applied d.c. voltage. Such a condition is more or less fulfilled in a magnetron once the threshold voltage V_0 necessary to maintain oscillations is reached, provided one considers the applied d.c. voltage to be $(V-V_0)$. In this application we are justified to some extent by the fact that the two experimental curves, so corrected, coincided exactly. The actual correction was made to the square root of the curves d and f, which, being curves of r-f power, are proportional to the stored energy in the circuit

[4] William Altar, private communication.

rather than the voltage. The corrected curve was then squared to give the solid curve of Fig. 4.

Since the r-f power out is at every instant proportional to the stored energy, the curves of Fig. 4 may be thought of as curves of stored energy (W_s) against time. We may use the curve of zero time of rise of the voltage pulse to determine the rate of change of stored energy in the build-up process. The rate of change of stored energy must be equal to the difference between the total power generated and the power lost both in the circuit and in the load. Since at present we are concerned with extending the data of Fig. 3, by the use of the curve of Fig. 4, we need not consider the circuit losses, for although they are directly proportional to the square of the slot voltage, as are the stored

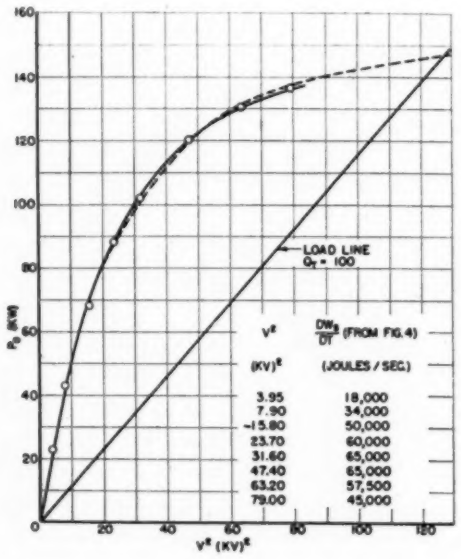


FIG. 5. Characteristic curve calculated from build-up curve.

energy, and the power out, each of these is independent of the others, and each may be considered an independent function of the square of the slot voltage. Thus we may consider the rate of change of stored energy equal to the difference between the net power generated and the load loss, where the net power generated is the total power generated minus the circuit losses.

The above relation may be expressed as follows:

$$\frac{dW_s}{dt}(V^2) = f(V^2) - \frac{\omega W_s(V^2)}{Q_T}, \qquad (2)$$

where W_s is stored energy, and $f(V^2)$ represents the net power generated as a function of the square of the slot voltage. Since the slot voltage completely determines the conversion process of the electrons, it will appropriately serve as independent variable. From the solid curve of Fig. 4 we obtain $dW_s/dt(V^2)$. To do this we measure the slope at various values of W_s and plot a curve of dW_s/dt against W_s, or what is the same thing, against V^2. So far the W_s scale of the curve of Fig. 4 has been undetermined; however, it will now be determined from the equilibrium point which is the intersection of the load line with the curve of net power generated. The load line is the plot of the last term of Eq. (2) which for a constant frequency and Q_T is obviously a straight line through the origin in a P_0 *vs.* V^2 plot (see Fig. 5). Now we have a portion of the curve of net power generated as shown in Fig. 3, and, dashed, in Fig. 5; for once the stored energy is built up, the net power generated is equal to the power out. The intersection point of these two curves gives the operating point of the magnetron in the steady state, and thus fixes the total amount of stored energy in the magnetron. Having chosen a Q_T and a frequency, we may calculate W_s from the V^2 of the intersection point. Using this final value of W_s we can now determine the scale of W_s in Fig. 4 so that this value is reached as a final W_s value of all the curves of Fig. 4.

Having thus determined the absolute (or actual) values of W_s in Fig. 4 and hence the values of dW_s/dt we may now extend the curve of Fig. 3 by the use of our $dW_s/dt(V^2)$ obtained from Fig. 4. First drawing our load line

$$P_0 = \frac{\omega W_s}{Q_T} = \frac{V^2}{Q_T z_0}$$

(Fig. 5) and adding to it our curve of $dW_s/dt(V^2)$ we obtain the points on the curve of Fig. 5, which is our desired net power generated. That is clear from Eq. (2)

$$f(V^2) = \frac{\omega W_s(V^2)}{Q_T} + \frac{dW_s}{dt}(V^2). \qquad (2a)$$

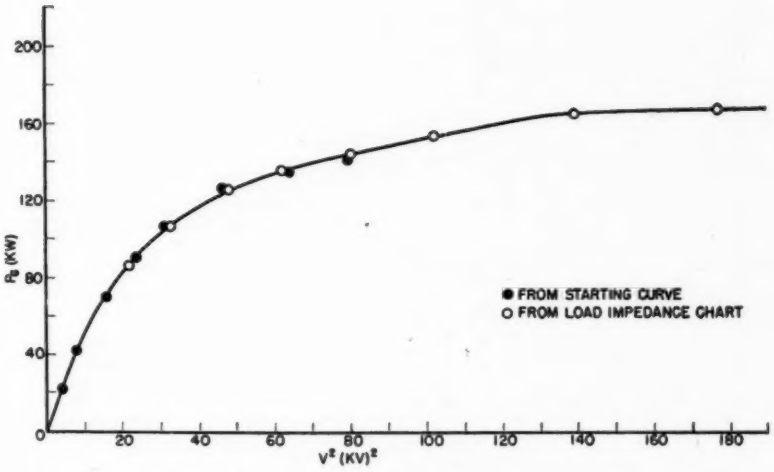


FIG. 6. Characteristic curve of 2J32 magnetron.

As stated above this curve is also the curve of P_0 *vs.* V^2 for steady-state operation. The curve of Fig. 3, obtained from the Rieke diagram is shown dashed in Fig. 5, and the solid curve is the curve obtained from the starting curve calculations. From the curve of Fig. 5 we may work backward and calculate starting curves for different load lines using the same graphical method described above. However, we shall first calculate the true curve of total power generated *vs.* slot voltage squared, and attempt to find an analytical expression for it before proceeding with any starting time calculations.

In order to determine the curve of total power generated *vs.* the square of the slot voltage, we must consider the circuit losses which we have so far found convenient and permissible to disregard. We will assume a constant unloaded Q_u for our magnetron as a preliminary step in the consideration of the circuit loss. The circuit efficiency now becomes

$$e=Q_u/(Q_u+Q_T)$$

and we may calculate the total power generated (P_g) from the output power (P_0) by using this circuit efficiency ($P_0=eP_g$). To do this numerically we must choose a numerical value for Q_u. Most S band magnetrons have unloaded Q's lying between 1100 and 1800. Using a value of 1500 we may correct the curve of P_0 *vs.* V^2 of Fig. 5 to get the curve of P_g *vs.* V^2 given in Fig. 6. From this curve it can be seen that the power generated approaches zero with a finite slope. This means that there is a limiting load beyond which the magnetron will not oscillate. This load corresponds to the load line which is tangent to the P_g curve at the origin, for this is obviously the last load line which can be drawn to intersect the curve of power generated.

Such a result is to be expected since it is clear that if the drain on the energy stored in the magnetron is so great that it effectively prevents the build-up of this energy above noise level, the magnetron will never be able to get started, and therefore will not oscillate. This is analogous to the conventional d.c. shunt generator in which the terminal voltage supplies both the exciting field and the load. There the power used by the load may so lower the terminal voltage that the exciting field will no longer be large enough to allow the generation of the power required by the load, and the generator will not operate.

Rather than continue with our graphical procedure we will attempt to fit an analytical expression to the curve of Fig. 6. This curve most obviously comes tangent to two lines, the limiting load line and the saturation power line. The limiting load line is a line through the origin with a finite slope, and the saturation power line is a straight line parallel to the V^2 axis. The simplest expression to fit such a curve is

$$f(V^2)=BV^2/(1+aV^2)=P_g, \qquad (3)$$

which, for $V=0$, has the finite slope B, and for large values of V^2 approaches asymptotically $P_g=B/a$. In order to test the fit of this function,

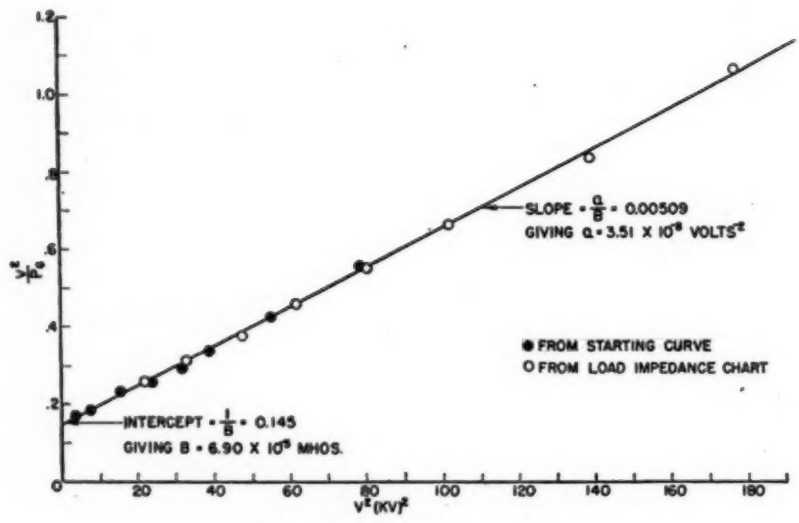

SLOPE = a/B = 0.00509
GIVING a = 3.51 X 10⁸ VOLTS⁻²
● FROM STARTING CURVE
○ FROM LOAD IMPEDANCE CHART
INTERCEPT = 1/B = 0.145
GIVING B = 6.90 X 10⁻³ MHOS.
V²/P₀
V² (KV)²

FIG. 7. Law of power generated.

we will write (3) in the form

$$V^2/P_g = 1/B + aV^2/B. \tag{4}$$

This expression is a straight line in the coordinates V^2/P_g and V^2 with an intercept $1/B$ and a slope of a/B. In Fig. 7 we see a plot of the data of Fig. 6 in the coordinates V^2/P_g *vs.* V^2. It is quite apparent that this plot is fitted by a straight line, which shows that the law proposed above does indeed represent the data of Fig. 6. This result is all the more remarkable when one considers that the data represented are from two different sources, the load impedance chart, and the starting curve. In Fig. 7 the data derived from the starting curve are represented by the closed circles while those derived from the load impedance chart are shown in the open circles.

SECTION II

Starting Time as a Function of Load

We now proceed to the integration of Eq. (2) using the $f(V^2)$ as found in Eq. (3). The result of this integration will then be solved explicitly for t and used to calculate the variation of starting time as a function of load.

If we now wish to include the circuit losses in Eq. (2), we find that it is best to combine them with the load and write, using Q_L instead of Q_T

$$\frac{dW_s}{dt}(V^2) = \frac{BV^2}{1+aV^2} - \frac{\omega W_s(V^2)}{Q_L}. \tag{5}$$

Assuming as is customary that W_s is the energy stored in the magnetron alone, we may write

$$W_s = A\,V^2,$$

where A is a constant equal to $1/\omega^2 L$ on our equivalent circuit basis. Equation (5) becomes

$$\frac{dW_s}{dt} = \frac{BW_s}{A+aW_s} - \frac{\omega W_s}{Q_L}. \tag{6}$$

Integrating we get

$$\frac{AQ_L}{Q_LB-\omega A}\ln(a\omega W_s) - \frac{Q_L}{\omega}\frac{Q_LB}{Q_LB-\omega A} \times \ln(Q_LB-\omega A-a\omega W_s) = t + \text{const.}$$

Now in order to evaluate the constant we must assume a value for W_s at $t=0$. This corresponds to the initial noise level from which the oscillations build up. This value of W_s shall be represented by (db) decibels below the final value of W_s. Now we write the final value of W_s (W_s final) as

$$W_s \text{ final} = (Q_LB-\omega A)/a\omega$$

for at this value of W_s, t goes to infinity. This leads to a constant of integration

$$\text{const.} = \frac{AQ_L}{Q_LB-\omega A}\{\ln(Q_LB-\omega A) - 0.23\text{ db}\} - \frac{Q_L}{\omega}\left\{\frac{Q_LB}{Q_LB-\omega A}\right\}\ln(Q_LB-\omega A).$$

Inserting this value of the constant of integration we get the final solution for t

$$t=\frac{AQ_L}{Q_LB-\omega A}\left\{\ln\left(\frac{a\omega W_s}{Q_LB-\omega A}\right)+0.23\text{ db}\right\}$$
$$-\frac{Q_L}{\omega}\left\{\frac{Q_LB}{Q_LB-\omega A}\right\}\ln\left(\frac{Q_LB-\omega A-a\omega W_s}{Q_LB-\omega A}\right). \quad (7)$$

If we define

$$1/Q=(B\omega L-1/Q_L),$$
$$\varphi=a\omega^2L/Q_L,$$
$$W=\varphi QW_s,$$

where $1/B\omega L$ is a negative input Q which represents the rate of energy input to the circuit, and $1/\varphi Q=W_s$ final, Eq. (7) now reduces to

$$t=\frac{Q}{\omega}\{\ln(W)+0.23\text{ db}\}-\frac{Q+Q_L}{\omega}\ln(1-W), \quad (8)$$

where the build-up curve is generated as W goes from zero to one.

Using Eq. (8), we may calculate the form of the build-up curve, and the starting time as a function of load and of noise level. Calculating first the starting time as a function of load we define the starting time as the time for W to reach the value 0.01. This is equivalent to W_s reaching 1 percent of its final value since for each load $W_s=W$ (W_s final). This definition of starting time agrees with that of More and Miley.[5] In Fig. 8a is a plot of the starting time, as defined above, against loaded $Q(Q_L)$. The curve was calculated from Eq. (8), using a noise level of 80 db down. The open circles are data of More and Miley on 2J32 tubes, and the closed circles are data on a 2J22. It appears that the agreement is poor in the region of light loading. It would seem, therefore, that we need to examine our assumptions more carefully. We derived Eq. (8) on the assumption that the stored energy in the magnetron was the only important stored energy in the system, and that it depended only on the square of the slot voltage and not upon the load ($W_s=AV^2$).

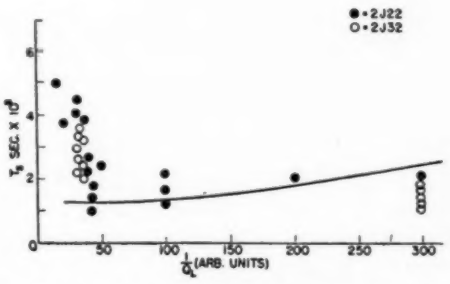


FIG. 8a. Starting time *vs.* load using stored energy in the magnetron alone.

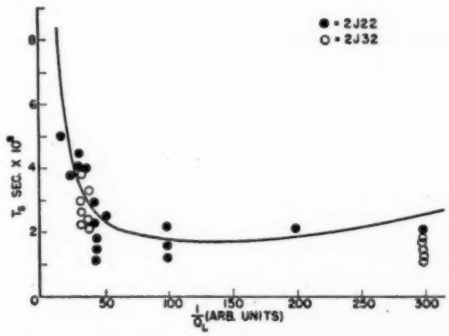


FIG. 8b. Starting time *vs.* load using total stored energy.

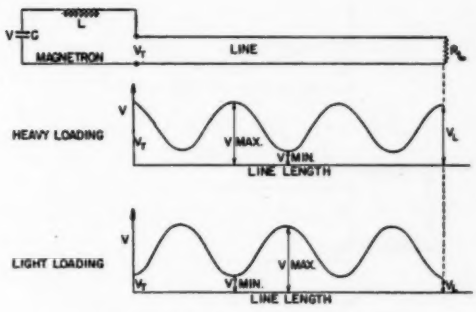


FIG. 9. Effect of load on V_T.

If we consider carefully the connection of the load to the stored energy, we will find that it exerts a remarkable influence on it. In Fig. 9 are shown the standing wave patterns for heavy and light loading. It is easy to see that the voltage V_t at the back of the magnetron cavity varies widely for these two cases. This means that the stored energy in the magnetron itself varies widely for these two cases even for essentially the same power output. In addition to this, there is some energy stored in the line when the line is not matched which adds a constant amount of stored energy to either case for the same *VSWR* and length of line. Taking these facts into consideration, we will correct Eq. (8) for the effect of load on the stored energy.

We shall first demonstrate the importance of stored energy contributed by the load, by means

[5] See reference 1, footnote 2.

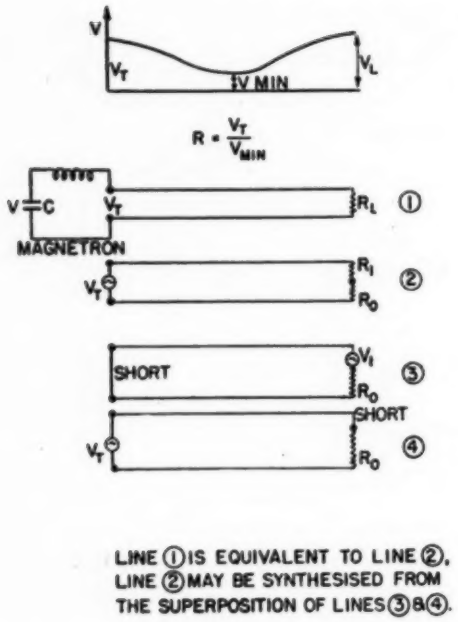


FIG. 10. Circuit for determining stored energy in line.

of the example of a shorted line section. It can be shown in general that if B is the susceptance of a resonant circuit the stored energy, E, may be written

$$E=\frac{V^2}{2}\frac{dB}{d\omega}\Big)_{B=0}. \tag{9a}$$

In the case of reactance (x) instead of susceptance we have

$$E=\frac{i^2}{2}\frac{dx}{d\omega}\Big)_{x=0}, \tag{9b}$$

where V is the voltage at the terminals where $dB/d\omega$ is measured, and i the current at the terminals where $dx/d\omega$ is measured. This is a corollary of Foster's reactance theorem. For a coaxial line or a wave guide we know that the terminal impedance Z is defined

$$Z=Z_0\frac{Z_L+jZ_0\tan(\omega l/c)}{Z_0+jZ_L\tan(\omega l/c)}$$

and the terminal admittance

$$Y=Y_0\frac{Y_L+jY_0\tan(\omega l/c)}{Y_0+jY_L\tan(\omega l/c)},$$

$$Y=jY_0\tan(\omega l/c) \quad \text{if} \quad Y_L=0.$$

From these expressions we may calculate

$$\frac{dB}{d\omega}\Big)_{B=0}=(l/c)\,Y_0\sec^2(\omega l/c)\Big)_{\omega l/c=0,\,\pi,\,2\pi,\,\text{etc}}$$
$$=(l/c)\,Y_0.$$

For

$$\omega l/c=0,\ \pi,\ 2\pi,\ \text{etc.}\ \ l/c=h\pi/\omega,$$

where h is the number of half-wave-lengths of line. Finally

$$\frac{dB}{d\omega}\Big)_{B=0}=\frac{h\pi Y_0}{\omega}. \tag{10a}$$

Similarly, we find

$$\frac{dx}{d\omega}\Big)_{x=0}=\frac{h\pi Z_0}{\omega}. \tag{10b}$$

If we consider the line of Fig. 10, and if we define (R) as the voltage at the terminals A divided by the voltage, one-quarter wave-length down the line, we may calculate the energy stored in the line as a function of the (R) and the terminal voltage V_t.

By the theorem of compensation we may consider the load R_L replaced by a load R_0 in series with a generator of zero internal impedance which generates such a voltage V_1, that its addition to the voltage V_0 gives the voltage drop through the original R_L. We will now consider the voltage in R_0 (matching resistance) as giving rise to energy in transit and the voltage V_1 as giving rise to stored energy in the line. First calculating the current i_1 through R_1 we will, by the theorem of superposition, replace the generator V_t by a short circuit (internal resistance of generator V_t=zero), then

$$i_1=V_1/R_0,$$

since the only generator now present is the V_1 generator. Now

$$V_1=V_tR_1/(R_1+R_0)$$

and

$$i_1=\frac{V_tR_1}{R_0(R_1+R_0)}=(V_t/R_0)(R-1)/R. \tag{11}$$

For $R<1$ we use the formula

$$W_l\ (\text{stored energy in the line})=\frac{i^2}{2}\frac{dx}{d\omega}\Big)_{x=0}$$

and obtain

$$W_l=\frac{V_t^2}{R_0^2}\left(\frac{R-1}{R}\right)^2(h\pi R_0/2\omega).$$

For $R>1$ we use formula (9a) giving

$$W_l=V_t^2\left(\frac{R-1}{R}\right)^2(h\pi G_0/2\omega),$$

which we see to be the same as for $R<1$ giving the stored energy in the line for any R

$$W_l=V_t^2\left(\frac{R-1}{R}\right)^2(h\pi G_0/2\omega). \qquad (12)$$

The energy stored in the magnetron (W_m) may be calculated from the transmitted $Q(Q_t)$.

$$Q_t=\frac{2\pi \text{ energy stored}}{\text{energy lost/cycle}}=\frac{2\pi W_m}{V_t^2G_L/f} \qquad (13)$$

or

$$W_m=V_t^2Q_tG_L/\omega.$$

Combining (12) and (13) to get the total stored energy, W_s, we have

$$W_s=W_l+W_m=\frac{V_t^2G_0}{\omega}\left\{RQ_t+\frac{h\pi}{2}\left(\frac{R-1}{R}\right)^2\right\}. \qquad (14)$$

In order to consider the variation of A, we must transfer the voltage V_t back to the slots of the magnetron. To do this we will use Eq. (1)

$$V^2=P_0Q_tz_0. \qquad (1)$$

Now the power output is

$$P_0=V_t^2G_L=V_t^2RG_0$$

for an even number of quarter-wave-lengths of line between the load and the terminals where V_t is measured.

For the magnetron $z_0=(L/C)^{\frac{1}{2}}$ enabling us to transform Eq. (1) into

$$V^2=V_t^2RG_0Q_t(L/C)^{\frac{1}{2}}$$

and the total stored energy, W_s, becomes

$$W_s=\frac{V^2g_0}{\omega}\left\{1+\frac{h\pi}{2Q_t^*}\left(\frac{R-1}{R}\right)^2\right\}, \qquad (15)$$

where $g_0=1/z_0=(C/L)^{\frac{1}{2}}$ for the magnetron and Q_t^* is the transmitted Q for matched line. Since by definition $W_s=AV^2$ we now see that A depends on the load, here represented by R.

$$A=g_0/\omega\left\{1+\frac{h\pi}{2Q_t^*}\left(\frac{R-1}{R}\right)^2\right\}. \qquad (16)$$

Examining Eq. (16) we see that for matched line ($R=1$). $A=g_0/\omega=1/\omega^2L$, as before. For $R>1$ (heavy loading) the term

$$(h\pi/2Q_t^*)\left(\frac{R-1}{R}\right)^2$$

approaches a constant finite value as R increases. Thus, the total stored energy approaches a finite value as the loads get heavier. Now as R decreases on the side of light loading, the term

$$(h\pi/2Q_t^*)\left(\frac{R-1}{R}\right)^2$$

goes to infinity, so that the total stored energy for the case of light loading approaches infinity as R goes to zero. This is quite reasonable since V_t must remain finite even when the *VSWR* goes to infinity, so that when V_t is a maximum of the *VSWR*, the total stored energy remains finite, but when V_t is a minimum of *VSWR* the

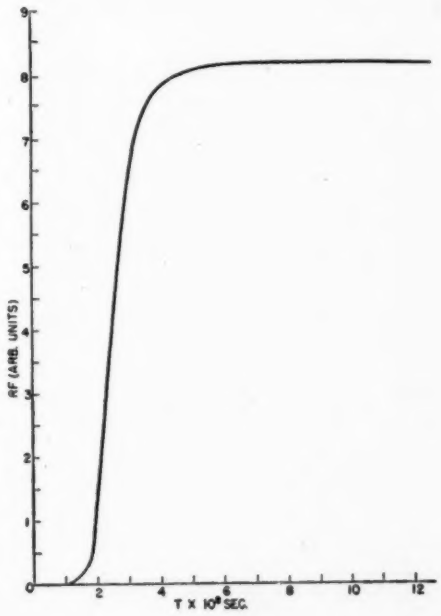


FIG. 11. Calculated build-up curve.

total stored energy should go to infinity as the (R) goes to zero.

Similarly, the loaded $Q(Q_L)$ used in Eq. (5) depends on the total stored energy,

$$Q_L=\frac{2\pi W_s}{(P_0+\text{loss})/f}=\omega W_s e/P_0, \qquad (17)$$

where e is the circuit efficiency. Using (16) and (17) in Eq. (7) we may calculate the starting time as a function of load with the stored energy in the line fully accounted for. The results of this computation are shown in Fig. 8b. Here the various constants used are $h=8$ half-wavelengths (a reasonable distance for a standing wave detector and a tuner) noise level -80 db down, $Q_t^*=100$. Here it is seen that the agreement between the calculated curves and the results of More and Miley (circles) is quite good —in the regions of both light and heavy loading.

SECTION III

Results and Conclusions

In Fig. 11 is given a curve of stored energy against time as calculated from Eq. (8). This curve may be compared with the solid curve of Fig. 4, for it is a build-up curve of the magnetron. It also refers to zero time of rise of the voltage pulse. The agreement is seen to be quite good.

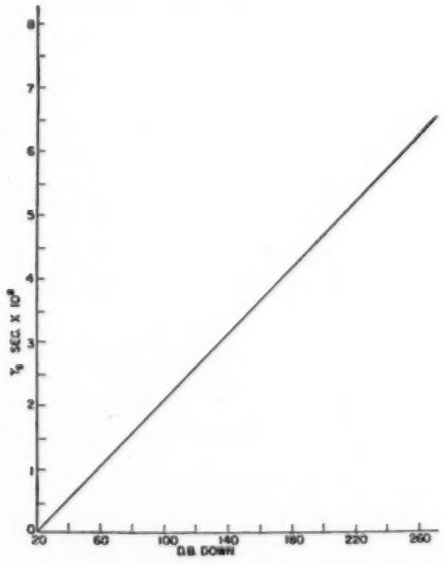


FIG. 12. Variation of starting time with noise level.

The slight difference of the top bends in the curves may be ascribed to the fact that the curve of Fig. 4 was corrected from experimental curves taken with voltage pulses that oscillated slightly upon reaching their peak values. The form of the build-up curve for different loads can be calculated by using Eq. (7) corrected by (16) and (17) but it does not seem to be worth while here.

In Fig. 12 is shown a curve of starting time against noise level for a matched load. This curve also was calculated from Eq. (8). It is clear that the noise level does not affect the starting time very much for any values which are physically possible. Noise levels from 160 db to 80 db down all give starting times of the right order of magnitude.

In this connection it was suggested by Dr. Altar that an undesirable mode could be suppressed by injecting an r-f voltage of the frequency of the desired mode, thus giving it a head start, so to speak, over the undesired mode. Such an experiment was performed by Dr. Coltman of this laboratory and was found to be successful.

In conclusion several points should be emphasized. First, it appears from the available data that a 2J32 magnetron has a limiting load beyond which it will not oscillate. This load appears to be in the neighborhood of a loaded Q of 15. At such a load the starting time becomes infinite so that the experiment of More and Miley should show a sharp increase of starting time if the curves were pursued to very low loaded Q values. Dr. Coltman has made magnetrons in the 3-cm band which were loaded to a Q_L of 12 with a frequency insensitive load, and found it quite impossible to get them to oscillate in the mode so loaded. The same tubes did oscillate when loaded to a Q_L of 20, showing that for a Q_L in the neighborhood of 15 oscillation ceased. This piece of evidence is cited in support of the result given above. Such evidence establishes with certainty the fact that the curve of Fig. 6 approaches zero with a finite slope.

Second, it should be pointed out that the curves of Fig. 6 and Fig. 7 are plotted from data taken with the d.c. voltage on the magnetron held constant, not the input power. This means that even though the power generated is a

monotonically increasing function of the r-f voltage on the slots, the conversion efficiency is not. In Fig. 13 is given the variation of conversion efficiency with the ratio of the r-f voltage on the slots of the magnetron to the applied d.c. voltage. It is seen that there is a maximum of conversion efficiency when $V_{r-f}/V_{d.c.}=0.65$. This result is in good agreement with the calculations of Dr. Altar[6] who gets such a maximum at $V_{r-f}/V_{d.c.}=0.55$, in his analysis of a bunching and catching mechanism as applied to energy conversion in magnetrons. The fact that the ratio obtained here is slightly larger than that of Altar is explained by the fact that the ratio of Altar is computed at a slight distance from the anode surface. This means that both the V_{r-f} and the $V_{d.c.}$ are somewhat reduced, but since the V_{r-f} falls off faster as one leaves the anode surface, the ratio should be somewhat less.

The variation of starting time with load, as shown in Fig. 8b, would appear to lend itself readily for use in "mode" suppression in magnetrons. The mode desired could be loaded in such a manner as to keep its starting time low while the same physical arrangement could be designed to load the undesirable mode in such a manner that it would have a much longer starting time.

[6] William Altar, private communication which will be published shortly.

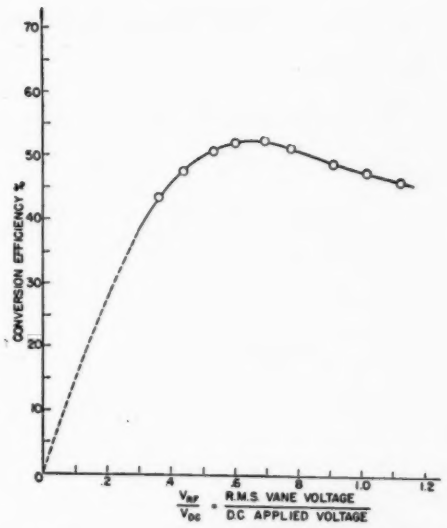


FIG. 13. Conversion efficiency *vs.* the ratio of slot to d.c. voltage.

This is possible because a line length can be chosen for two modes of different frequency so that one will have a standing wave pattern in the line similar to that shown in Fig. 9 for "heavy loading" and the other will have a pattern similar to that shown for "light loading."

The author wishes to acknowledge the assistance of Dr. John W. Coltman without whose constant advice this paper would not have been completed.

The Measured Impedance of Cylindrical Dipoles

D. D. KING

Cruft Laboratory and Research Laboratory of Physics, Harvard University, Cambridge, Massachusetts

(Received April 2, 1946)

The input impedance of cylindrical dipoles is considered from the experimental point of view. Errors in measuring methods are considered and the results of a recent determination of dipole impedance are displayed.

INTRODUCTION

IN recent years considerable advance has been made in the theory of antennas of various shapes. The theoretical treatments regularly include an approximation of the cylindrical dipole. The purpose of this paper is to provide experimental verification of the results obtained by mathematical solutions of the problem.

The significance of experimental results depends upon an adequate definition of the quantity being measured as well as upon the precision of the experimental technique. Differences between theoretical and physical boundary conditions are considered briefly in this connection. The experimental methods are described and their accuracy is specified. With this information the results are useful in the interpretation of other data on antenna input impedance. The measured values displayed in this paper have already been quoted to this end elsewhere.[1,2]

PHYSICAL BOUNDARY CONDITIONS

Analytical solutions for the cylindrical dipole must satisfy Maxwell's equations and the prescribed boundary conditions. In general certain simplifications in the boundary conditions are required to obtain a mathematical solution to this problem.[2,3] The discrepancy between theory and experiment attributable to differences in the boundary values may be minimized by careful experimental methods as well as by refinements in analysis. The dipole model used for the present measurements represents certain boundary values which are now considered.

FIG. 1. Currents at the driving point of a half-dipole and image plane.

The antenna configuration consisted of a half-dipole and image plane. For measuring purposes the dipole is mounted on the inner conductor of a concentric line the outer conductor of which is attached to the image plane. The coaxial line-spacing produces an annular gap at the center of the image plane. The antenna is driven by the fields in this region and hence the input impedance is also defined here. The currents at the driving point are indicated in Fig. 1.

There would seem to be a phase lag between the actual currents on the image plane and those which would exist in the ideal case of a base-driven dipole. The amount of lag is $\beta(b-a)$, where a and b are the inner and outer radii of the line and $\beta=2\pi/\lambda$. The magnitude of the quantity $\beta(b-a)$ is determined by other considerations to be examined later. Thus, a principal consequence of the concentric line-spacing is to introduce a lag in the currents in the image antenna proportional to this quantity. Discontinuities at the junction of line and antenna conductors may cause an added lumped capacitance. The geometry at the junction of line and antenna shown in Figs. 2a and 2c applies to the present measurements. Arrangements shown in Fig. 2b were not used.

Another consequence of the finite spacing

[1] R. King and D. Middleton, Quart. App. Math. **3** (Jan. 1946).
[2] D. Middleton and R. King, J. App. Phys. **17**, 273 (1946).
[3] L. Brillouin, Elec. Commun. **21**, 4 and 22 (1943).

$(b-a)$ is to distribute the driving field along a segment of the center conductor of approximate length $(b-a)$. The physical boundary condition for concentric-line excitation therefore does not contain well-defined terminals; instead it involves a potential difference distributed about the base of the half-dipole. Thus the actual generator to be approximated by theoretical analysis is in the form of a thick circular ring, or toroid, which maintains a potential difference along a continuous segment of conductor.

Impedance measured between terminals with appreciable spacing is of course not precisely defined. The fundamental relation for scalar and vector potentials in the interval AB is the following

$$\int_A^B \mathbf{E}\cdot ds = -\int_A^B \left\{ j\omega\mathbf{A} + \frac{d\phi}{ds} \right\} ds. \qquad (1)$$

In case the interval between the terminals A and B is very short, the term in the vector potential $\mathbf{A}$ may be neglected. The integral then becomes independent of the path AB. The usual definition of impedance in terms of the scalar potential difference may then be formulated

$$Z_{AB} = (\phi_A - \phi_B)/I. \qquad (2)$$

Evidently a slight error is always present at the driving point due to the finite line-spacing. In an exact comparison with theory this relatively minor source of error may be properly considered. Factors that prevent an indefinite reduction of the spacing at the terminals are examined in the next section.

CHOICE OF DIMENSIONS

The scale of the apparatus and antennas used depends on the choice of wave-length. In making this choice, two basic limitations are encountered. At long wave-lengths the earth introduces errors since it is an imperfect conductor and not well suited for use as an image plane. On the other hand, extremely short wave-lengths decrease the inherent accuracy of the apparatus. This is because dimensions and tolerances decrease in proportion to the wave-length. For operation in the laboratory the range between 10 cm and 100 cm is indicated. In this region an accuracy of 0.1 mm in position is adequate and readily attainable, as are the usual construction tolerances of ± 0.001 inch. An image surface of ample proportions may also be provided without resort to a field station.

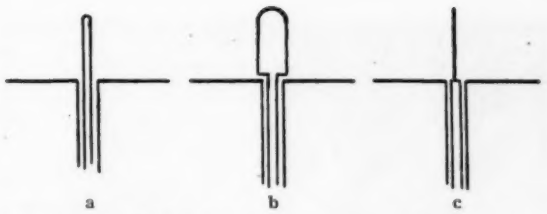


FIG. 2. Half-dipole and coaxial line configuration.

The choice of characteristic resistance of the measuring line is dictated by the loads to be measured. In general, the best accuracy is achieved when power is efficiently transferred to the load. However, since impedance measurement often involves the use of cables, the available fittings and cables must be considered. In this instance a line resistance of 49 ohms was chosen.

Having thus fixed the ratio of conductor diameters, the magnitude of the quantity $(b-a)$ at the image surface is fixed by the antenna diameters to be used. This precludes very small values of $(b-a)$. The small dimensions of the line at this point are not favorable to measuring procedures since closer tolerances are required in a thin line. Accordingly, a 16-inch linear taper was used in the present measurements which everywhere maintained the characteristic resistance. No reflections from this tapered section were detectable and the measuring line diameter could thereby be set at 1 inch while the hole in the image plane was only $\frac{5}{16}$ inch, which gives $\beta(b-a) = 0.05$ rad. at the standard wave-length of 28.00 cm. The phase lag introduced by the hole in the image plane, though small, is therefore not entirely negligible.

MEASURING METHODS

In making antenna impedance measurements it is possible to choose one of two common parameters. Most theoretical treatments of the subject use the ratio of antenna length to diameter as the parameter. This is commonly written in the form of $\Omega = \log_e(2h/a)$ where h is the half-length and a the radius of the antenna. From the experimental point of view this implies

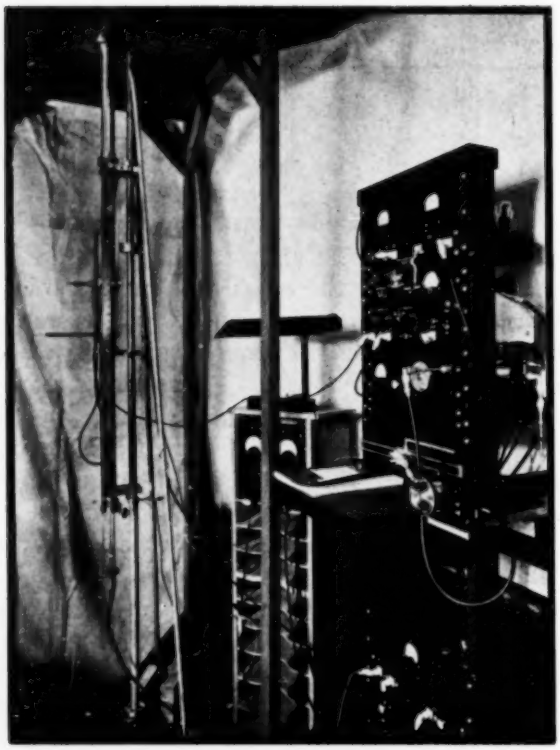


FIG. 3. Measuring apparatus: the line is at the left; monitors, amplifiers, and oscillator at the right.

varying the frequency. In order to include all lengths from the half-wave-length down to very short antennas, a very wide range of frequencies must be used. This is troublesome since the errors introduced by the finite image plane will change. Moreover, the tuning of the measuring equipment and the detector calibration all require attention as the frequency is varied over a wide range. Accordingly, another parameter was chosen in which a/λ is held constant and the half-length h changed. Since the theoretical parameter Ω is logarithmic, and hence slowly varying, the discrepancy between theory and experiment is negligible in any small interval. By measuring various thicknesses a curve with constant Ω can be obtained. The experimental procedure involves shortening the antenna for each reading. This is hardly more time-consuming than the readjustment of the apparatus would be, and permits greater accuracy in the results for the reason mentioned.

In order to assure maximum accuracy it seemed advisable not to rely exclusively on the standard standing-wave-ratio (SWR) method of measurement. For small losses in line and load this method is open to criticism because of the possibility of detector loading. An apparatus permitting both the conventional SWR method and the resonance curve method to be used provides greater flexibility in obtaining results for all values of terminal impedance. The use of resonance curves implies a movable piston at one end of the line, and this eliminates the possibility of using solid dielectric filled lines. The presence of dielectric beads, strips, or pins in the line is highly undesirable because of the discontinuities and reflections occasioned by them. It is possible to calculate the position of two beads with perfect dielectric which will produce no reflections. However, the calculation in general demands that the terminal impedance be known.

In order to avoid the errors due to movable beads and other types of dielectric supports, the measuring line used for the present measurements was designed to avoid the use of all supports. This was accomplished by mounting the conductors vertically. Several important advantages accrue from this type of mounting. Thus, the image plane could be located several feet above the operator, thereby minimizing disturbing effects from objects in the room. The weight of the inner conductor and attached antenna is borne by the brass base-plate at the bottom of the line. At the extreme upper end of the line a thin, perforated polystyrene wafer is inserted flush with the surface of the image plane. This wafer, being less than 0.004λ thick (1 mm), may be treated as a lumped capacitance in parallel with the load. The reactance introduced by the polystyrene may be measured and removed from the final results by subtracting its admittance from the load admittance. The fringing, or throat effect, at the open end of the line is probably also somewhat compensated by this procedure.

The wafer support maintains the line spacing well since gravitational effects are eliminated. The space between the two coaxial conductors is then left free to permit a piston to travel up and down for resonance curve measurements. Adequate rigidity is obtained from the silvered ground steel (drill rod) inner conductor and the $\frac{1}{4}$-inch wall of the aluminum outer conductor.

Tolerances were better than ± 0.001 inch. The radii of the inner and outer conductors were $a = 0.5588$ cm and $b = 1.2738$ cm. The calculated characteristic resistance is $R_c = 49.38$ ohms; the attenuation constant $\alpha = 3.78 \times 10^{-4}$ neper/cm. At the standard wave-length of 28.00 cm, the phase constant was $\beta = 0.2244$ radian/cm.

A movable tuned probe is provided for detecting the voltage amplitude along the line; probe depth is readily adjustable by a lead screw. The output of a Sylvania 1N21B crystal in the probe is fed through a selective amplifier into a Ballantine voltmeter. The voltage gain of this amplifier is 78 db and the noise level about 78 db below one millivolt. Extensive grounding proved to be essential in maintaining the noise level. Power is furnished by a 2C40 triode in a tuned cavity; 1100-cycle square-wave modulation is applied. The measuring line and associated equipment is shown in operating condition in Fig. 3.

MEASURING PROCEDURE

For low values of damping the resonance-curve method of measurement was used; for high damping, the standing wave ratio supplemented the resonance curve data. Circuit adjustments to permit changing from one method to another could be made quickly. Particular emphasis was placed upon the monitoring in order to eliminate fluctuation errors.

In resonance curve measurements it is essential that the oscillator be loosely coupled to the line. This is accomplished by the circuit shown in Fig. 4. A concentric-line-type wave meter and a bolometer amplitude-monitor are continuously coupled to the circuit. Constant information on the output amplitude and frequency is furnished by direct reading meters. Amplifiers provide a

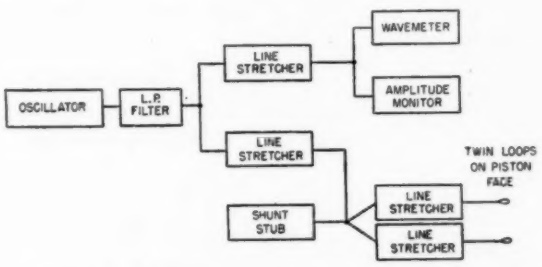


FIG. 4. Block diagram of the high frequency circuit.

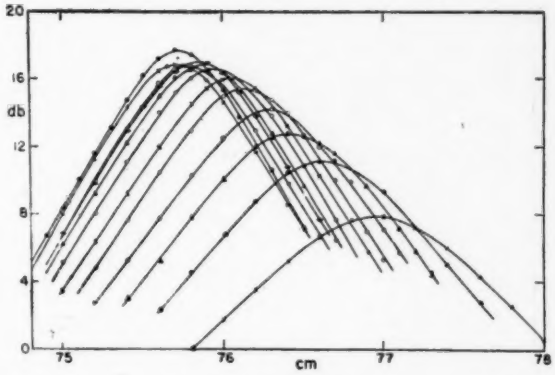


FIG. 5. Original data in the form of resonance curves. The wave-length is 28 cm and the width was taken at a level 3.16 db below the peak of each curve.

monitor sensitivity equal to the sensitivity of the detector on the line. If the amplitude- and frequency-monitors are undisturbed by the tuning of the measuring line through the required interval, then the coupling between oscillator and measuring line is loose, *by definition*.

In order to provide flexibility in adjustment two separate coaxial feeders are provided in the piston assembly. Each feeder terminates in a loop 3 cm long on the face of the upper piston. Tandem pistons with 3-cm brass flutes assure complete decoupling between the line and the cavity located between the back of the pistons and the line base-plate. Each feeder cable is attached to an independent phasing stub in a twin line-stretcher. A shunt stub is placed to form a four-way junction with these feeders and the line from the oscillator. This shunt stub forms the mutual impedance Z_{12} between the oscillator (primary) and measuring line (secondary). Suitable stops on the stub permit rapid adjustment of the coupling between oscillator and line.

For standing-wave-ratio measurements a larger amplitude is required along the line, since the minima of the distribution curve must be measured. The resonance curve procedure deals only with the peak of the resonance curve and hence requires less power input to obtain adequate deflections. Any desired degree of coupling can be obtained readily by means of the shunt stub mentioned above. Provision is made to permit accurate reproduction of the settings of this stub. This is needed in the resonance curve procedure, which requires the

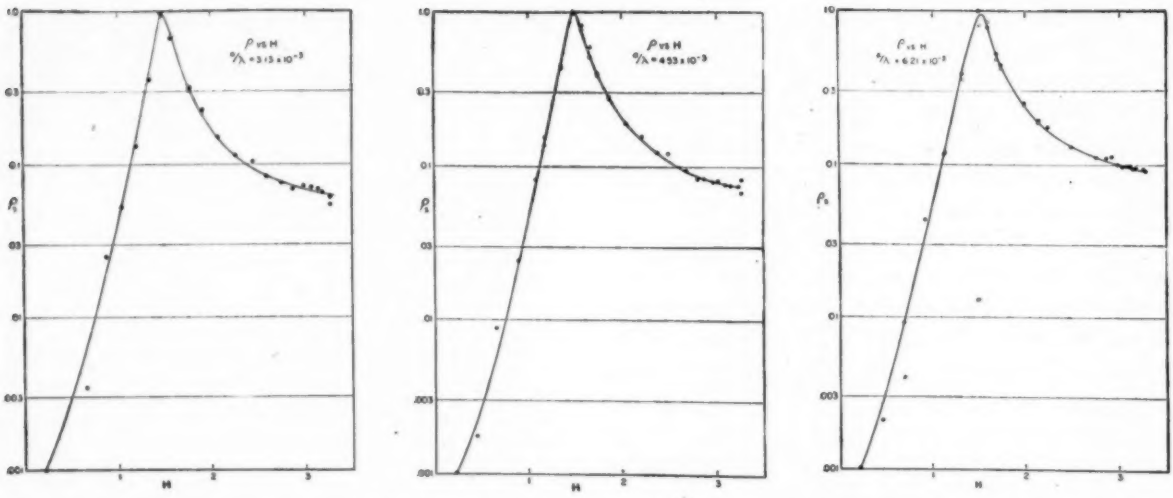


FIG. 6. Termination damping ρ in nepers as a function of $H=\beta h$, the dipole half-length in radians; the radius of the dipole is a. Solid points are from resonance data; circles from standing-wave data.

terminal impedance at the sending end to remain constant between comparisons with the standard termination.

An auxiliary probe near the sending end of the slotted section of line may be inserted to permit a direct test of probe loading. Probe effects cancel out in the resonance curve procedure, but the standing-wave method may suffer from probe loading effects at higher standing-wave ratios.[4] No loading effects could be detected for the standing-wave ratios measured.

ACCURACY OF MEASUREMENTS

The accuracy of which the apparatus is capable depends principally on two factors: first, the precision of linear measure along the line; and second, the accuracy of the detector reading. Steel scales and verniers permit readings on the present apparatus to within 0.1 mm. Since a sliding contact is involved, some irregularities might be expected. However, the original data yielded uniformly smooth resonance curves. A sample family of resonance curves for an antenna load is shown in Fig. 5. A fair estimate of the accuracy of the data on the position of the curves is probably better than ± 0.1 mm or $\pm 0.0004\lambda$. This is because the aggregate of all the points determines this position, which is therefore more accurately known than the location of any one point. Tests on the standard copper-plate termination made at regular intervals bear this out. The width of the curves is given to nearly the same precision, but additional error is introduced here in fixing the level at which the width is to be measured.

The width of the curves was measured at $(1/\sqrt{2})V_{\max}$. This level, somewhat higher than the usual half-power point, was chosen to extend the range of the method in the direction of high damping (low SWR), thereby securing a greater overlap with the standing-wave-ratio method. Increased accuracy in measuring the width of the curves is required at the level chosen, but the limit of error mentioned above is sufficiently small to permit this. The square of the amplitude level was read on a Ballantine voltmeter. An accuracy within ± 0.1 db was assumed. Repeated calibration tests were made and the deviations on the calibration curve confirm this as the likely error.

The original experimental results are in terms of termination phase shift Φ and termination damping ρ.[5,6] The resistance and reactance of the load are functions of both these independent variables.

$$R=2R_c\frac{\sinh 2\rho}{\cosh 2\rho-\cos 2\Phi}, \tag{3}$$

$$X=-2R_c\frac{\sin 2\Phi}{\cosh 2\rho-\cos 2\Phi}. \tag{4}$$

[4] Altar, Marshall, and Hunter, Proc. I.R.E. **34**, 33 (1946).

[5] R. King, J. App. Phys. **14**, 577 (1943).

[6] R. King, Proc. I.R.E. **29**, 640 (1941).

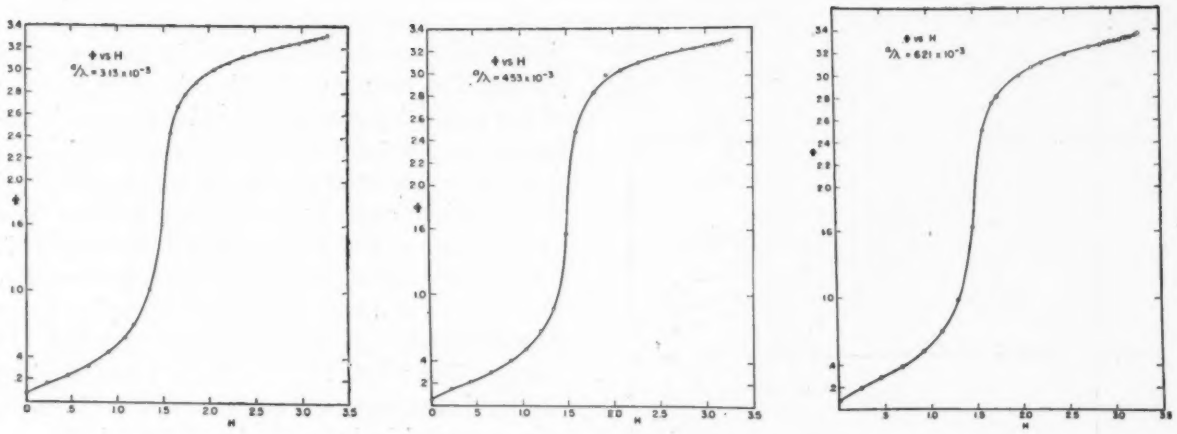


FIG. 7. Termination phase shift in radians as a function of the dipole half-length in radians.

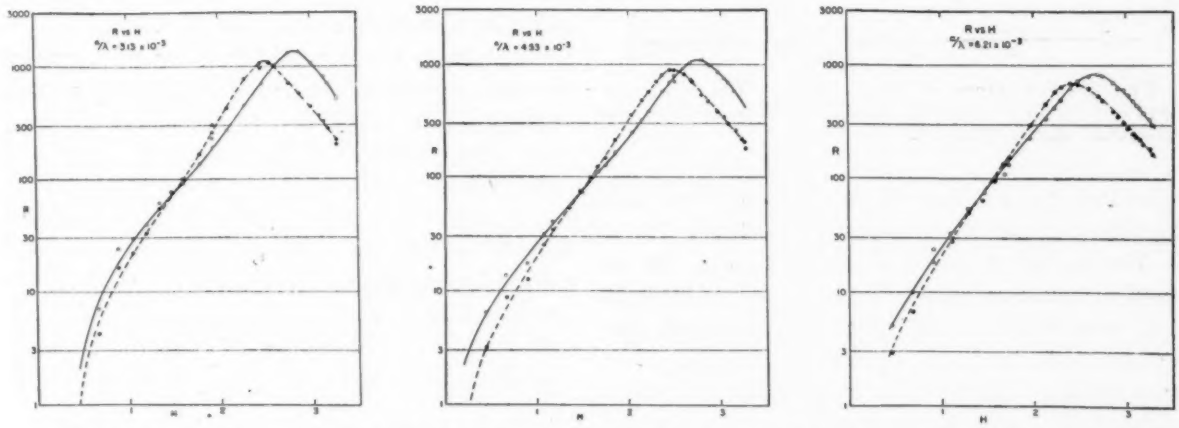


FIG. 8. Input resistance of dipole of radius a as a function of its half-length in radians. The solid curve has been compensated for the shunting effect of the polystyrene support; the dashed curve has not.

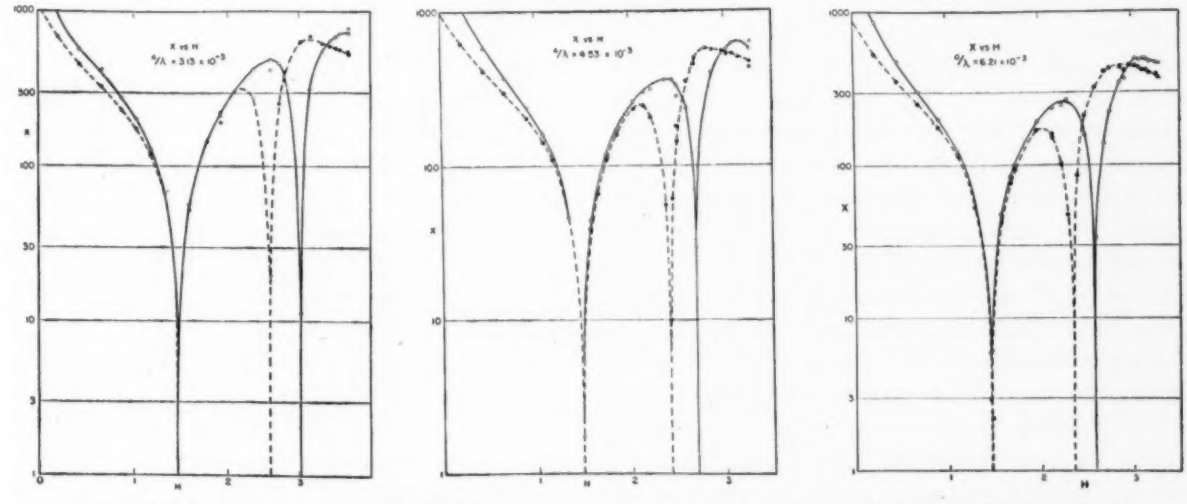


FIG. 9. Input reactance of a dipole of radius a as a function of its half-length in radians. The solid curve has been compensated for the shunting effect of the polystyrene support; the dashed curve has not. The sign is not contained in the logarithmic plot.

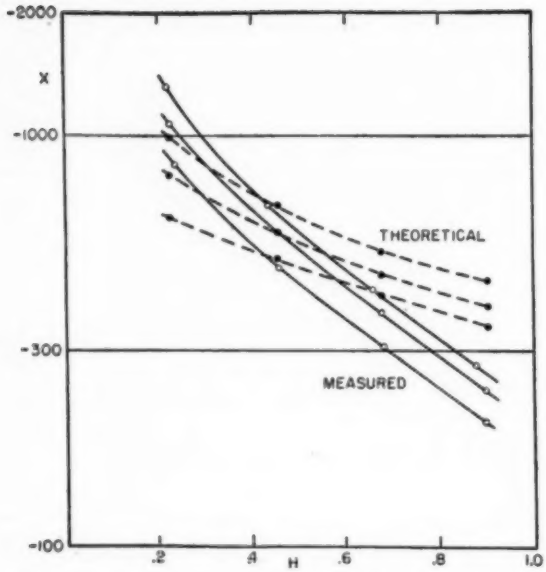


FIG. 10. Reactance of short dipoles. The measured curves are taken from Fig. 9 and the theoretical values calculated electrostatically.

The factor 2 in Eqs. (3) and (4) converts the results to symmetrical dipole impedance.

The accuracy of the ohmic results is not so easily stated, since the transformation from Φ and ρ to R and X may increase or reduce the inaccuracies in the measured variables. A measurement depending principally on the phase-shift function Φ was made at regular intervals to determine the impedance of the open end with associated wafer support. A value of $-j1440$ ohms ± 5 percent was obtained. Since polystyrene is somewhat affected by the humidity, some of the deviation recorded is probably not caused by the measuring technique.

Measurements made at large terminal damping depend principally on the damping function ρ. Independent determinations of ρ were made, using the resonance-curve method and the standing-wave-ratio method. The two methods agreed very closely. The experimental points from both methods are plotted in the curves. An error of the order of 5 percent is indicated by the location of the points along the curves for ρ. Values taken from the smooth curve should be more accurate. A much smaller deviation of the experimental points is apparent in the curves of the phase-shift function Φ. The curves are shown in Figs. 6 and 7.

The curves of R and X shown in Figs. 8 and 9 show both compensated and uncompensated values. The compensation involves elimination of the parallel reactance of -1440 ohms introduced by the polystyrene wafer support. Evidently the presence of even a small section of dielectric influences the results appreciably. The significance of conditions at the driving point is therefore borne out by the experimental results. The curves for ρ and Φ are slowly varying when the corresponding R and X curves exhibit rapid changes. Therefore the use of the ρ and Φ plots effectively increases the number of experimental points available in critical regions of the resistance and reactance graphs. Several additional points were chosen in this manner about the resonant and anti-resonant lengths.

The image plane used consisted of a square sheet of aluminum 90 cm on a side, supplemented by eight brass rods interconnected by heavy wire. The total area of this assembly was about eight times that of the center section. Data taken with only the inner section in place differed from values taken with the added section by less than 3 percent for several antenna lengths. The phase function Φ remained unaltered, while the damping ρ increased slightly with the size of the plane. A 5 percent increase in

TABLE I. Critical values of cylindrical antenna, $\Omega = 10$.

	Experimental D. D. King	King-Middleton (second order)
$(R_0)_{\text{anti-res}}$	820	860
$\pi - \beta h_{\text{anti-res}}$	0.60	0.614
$\lvert X\rvert_{\min}/\lvert X\rvert_{\max}$	1.95	1.8
$(R_0)_{\text{res}}$	71.5	71
$\pi/2 - \beta h_{\text{res}}$	0.098	0.094
$(R_0)_{\beta h = \pi/2}$	85	88
$(X_0)_{\beta h = \pi/2}$	47	42.5

TABLE II.

	D. D. King compensated	D. D. King uncompensated	Brown and Woodward	$a/\lambda \times 10^{-3}$
$(R_0)_{\text{anti-res}}$	1320	1080	1340	3.13
	1080	880	1050	4.53
	820	700	800	6.21
$\pi - \beta h_{\text{anti-res}}$	0.36	0.70	0.47	3.13
	0.44	0.72	0.59	4.53
	0.60	0.79	0.68	6.21
$\lvert X\rvert_{\min}/\lvert X\rvert_{\max}$	1.55	2.0	1.8	3.13
	1.88	2.32	2.3	4.53
	1.95	2.57	4.2	6.21

damping was recorded for flat-topped antennas as compared with round-topped types of equal length. Again the phase function Φ remained constant. The data available at present are inadequate to permit conclusions from these special results pertaining to the image plane and the end-surfaces. Further measurements are indicated to elucidate the effect of the end-surfaces and of the image plane.

COMPARISON OF RESULTS WITH THEORY AND WITH OTHER MEASUREMENTS

Several critical points in the impedance characteristics of an antenna may be chosen for comparison of the experimental results obtained with calculated values. In general, the maximum, minimum, and zero values of reactance are determined primarily by phase-shift data, while the resistance is also a function of the damping. In view of the greater precision of phase-shift determinations, somewhat greater accuracy may be expected in the reactance data than in the resistance values. Table I shows the results of a recent theoretical treatment and of the present measurements. The theoretical aspects of the numerical results have been discussed with a previous publication of this table.[2]

A plot of certain critical values of resistance

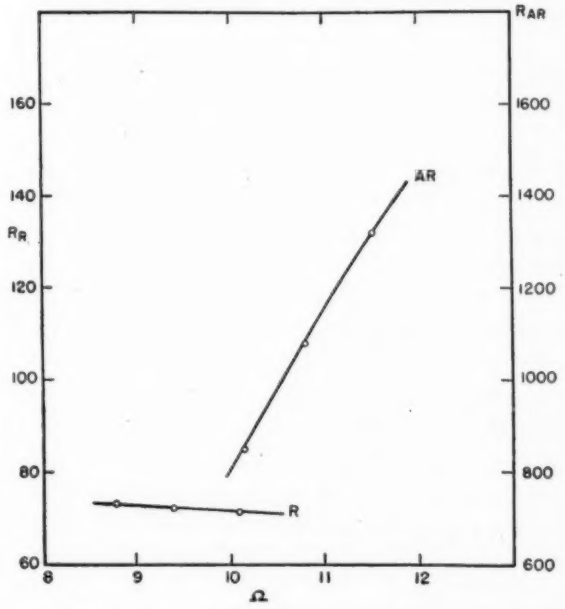


FIG. 11. Resonant and anti-resonant resistance as a function of the thickness parameter $\Omega = 2 \ln (2h/a)$.

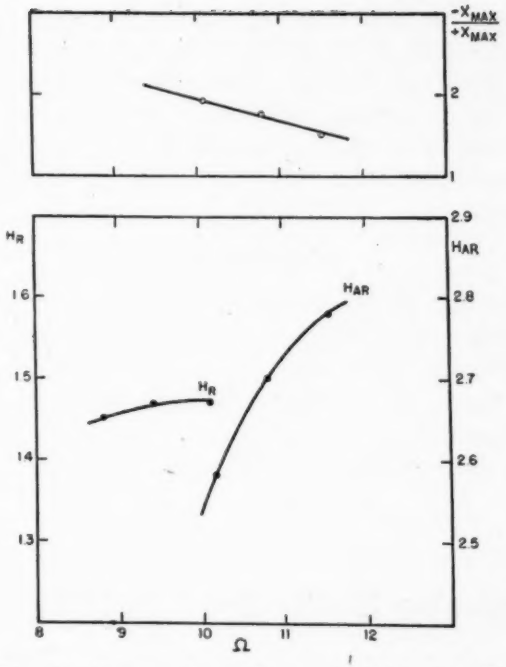


FIG. 12. Top: ratio of the magnitude of maximum negative and positive reactance; bottom: resonant and anti-resonant half-length in radians, all as a function of the thickness parameter $\Omega = 2 \ln (2h/a)$.

and reactance is shown in Figs. 10, 11, and 12. The curves of the reactance of very short dipoles are of particular interest. In case the antenna is short ($\beta h = H < 1$), an electrostatic calculation of the capacitance can be made[7] and X set equal to $-1/\omega C$. A very reliable theoretical result is thereby obtained. Agreement between theory and experiment here appears good. In case of extremely short antennas, the measurement is no longer accurate since the hole in the image plane and the antenna length are of the same order of magnitude. Likewise, antennas longer than $H = 0.5$ radian no longer admit use of a lumped capacitance determined from an electrostatic analysis.

A comparison with experimental results of Brown and Woodward is included in Table II. These writers used a 5-meter wave-length with a conventional slotted line. The agreement is rather good, with the exception of the reactance ratio. This is to be expected, however, since Brown and Woodward[8] directed their measure-

[7] R. King and C. W. Harrison, J. App. Phys. **15**, 170 (1944).

[8] G. H. Brown and O. M. Woodward, Proc. I.R.E. **33**, 257 (1945).

ments toward cases of practical engineering significance. Considerable base capacitance is encountered in base-driven dipoles attached through standard cable connectors. The effect of base capacitance is clearly indicated by the uncompensated results of the present measurements.

CONCLUSION

Graphs of the measured impedance characteristics of cylindrical dipoles have been presented. Separate plots for various thicknesses are intended to display the individual experimental points clearly. The results are given in this fashion together with specifications of the experimental procedure and accuracy in order to facilitate the correlation between antenna theory and experiment. Although the available apparatus cannot conform in full to the theoretical models described, a close approximation was generally achieved.

The writer is indebted to Professor R. W. P. King for many valuable suggestions in connection with the measurements described. These results are incorporated in a portion of the writer's Ph.D. thesis at Harvard University. The apparatus involved was constructed at Central Communications Research, Cruft Laboratory, under the auspices of the Office of Scientific Research and Development.

Elementary Treatment of Longitudinal Debunching in a Velocity Modulation System

EUGENE FEENBERG
Washington Square College, New York University, New York, New York
(Received May 14, 1946)

Certain configurations of anode, grid sections, and uniformly emitting cathode produce rectilinear motion of the electrons, both with and without velocity modulation. These systems are (a) plane parallel electrodes and grid sections of unlimited extent, (b) coaxial circular electrodes and grid sections of infinite length along the axis, and (c) concentric spherical electrodes and grid sections. The uni-dimensional character of the motion permits a complete evaluation of the effect of space charge on the bunching process in the range of drift space where overtaking (crossing of orbits) does not occur.

THE theory of velocity modulation has been developed from two distinct points of view exemplified in D. L. Webster's kinematic analysis of the bunching process[1] and Hahn, Metcalf, and Ramo's treatment of the electron beam as a medium for the propagation of electromagnetic disturbances.[2] There exist problems in which both points of view can be combined to advantage, retaining the rigorous foundation of the second and also a large measure of the practical usefulness of the first. It is the purpose of the current note to present a unified treatment of three closely related problems of this type.

Certain configurations of anode, grid sections, and uniformly emitting cathode produce rectilinear motion of the electrons, both with and without velocity modulation. These systems are (a) plane parallel electrodes and grid sections of unlimited extent,[3] (b) coaxial circular electrodes and grid sections of infinite length along the axis, and (c) concentric spherical electrodes and grid sections.

The uni-dimensional character of the motion permits a complete evaluation of the effect of space charge on the bunching process, subject to the following assumptions and limitations:

(1) In the space between buncher and catcher grid sections the negative space charge produced by the steady electron current is exactly neutralized by a positive ion space charge. In order to secure a precise mathematical problem it may

[1] D. L. Webster, J. App. Phys. **10**, 501 (1939); R. H. Varian and S. F. Varian, J. App. Phys. **10**, 321 (1939).
[2] W. C. Hahn and G. F. Metcalf, Proc. I.R.E. **27**, 106 (1939); W. Ramo, *ibid.* **27**, 757 (1939); W. C. Hahn, Gen. Elec. Rev. **42**, 258 (1939), **42**, 497 (1930).

[3] Space charge effects in system *a* are discussed in reference 1 and also in unpublished notes by W. W. Hansen.

be assumed that the positive ions move with the same velocity as the incoming electrons.[2] There is then no steady field (neither electrostatic nor magnetostatic) acting upon the beam.

(2) The ratio of beam velocity to the velocity of light is a small quantity ($v_0/c \ll 1$).

(3) The ratio of buncher voltage to beam voltage is a small quantity ($\alpha = V_1/V_0 \ll 1$). Here the form $V_1 \sin \omega t_1$ is taken for the instantaneous voltage across the buncher grids.

Velocity modulation at the buncher causes hf deviations from the steady-state electron distribution. These deviations produce a hf component of electric field parallel to the direction of propagation. The field in turn modifies the bunching process. We seek a self-consistent solution of the system of field equations and equations of motion in the sense that the field together with the velocity modulation at the buncher produces exactly the motion required to give rise to the field. In the theory of atomic structure problems of this general character are solved by methods of successive approximation (Hartree, Hartree-Fock equations). Fortunately a simple analytic, solution exists in the present case.

Let z represent the distance measured along the direction of propagation (with the origin at the axis or center of symmetry in systems *b* and *c*). In the usual notation $\tau = \omega t_1$ is the phase at which an electron passes the effective center z of the buncher grid space and $\theta + \omega(z - z_1)/v_0 = \omega t$ is the phase at which the same electron reaches the position z. Suitable independent variables are provided by z and t or by z and θ. Thus any hf quantity can be expressed as a function, $f(z, \theta)$, of z and θ or as a function, $f'(z, t) \equiv f(z, \theta)$, of z and t.

The boundary condition which correlates the field in the drift space with the amount of velocity modulation is furnished by the equation

$$\dot{z} = v_0[1 + \tfrac{1}{2}\mu\alpha \sin \tau], \quad z = z_1. \tag{1}$$

Here v_0 is the beam velocity, μ the beam coupling coefficient,[4] and α the ratio of buncher voltage to beam voltage.

[4] For system *a*

$$\mu = \sin(\tfrac{1}{2}s)/\tfrac{1}{2}s$$

in terms of the transit angle s through the buncher grid space. The same relation holds for systems *b* and *c* if the buncher grids are closely spaced in comparison with the mean distance of the grids from the axis or center of symmetry.

A point of departure for the calculation is supplied by the tentative hypothesis that the relation between θ and τ is formally the same as in the kinematic bunching theory. Now the bunching parameter appears as an unknown function of z in the cycloidal equation[1]

$$\theta = \tau - X(z) \sin \tau. \tag{2}$$

The condition $X(z) \leqq 1$ holds throughout the following discussion. This restriction is necessary and sufficient to exclude crossing of orbits and allows the hf component of current density to be computed from the relation

$$i(z, \theta) = i_0\left(\left(\frac{d\tau}{d\theta}\right)_z - 1\right)\left(\frac{z_1}{z}\right)^n, \tag{3}$$

in which i_0 is the entering current density at the center of the buncher grid section. The integer n has the values 0, 1, and 2 for systems *a*, *b*, and *c*, respectively. Equation (3) expresses the conservation of charge if no overtaking (or crossing of orbits) occurs in the beam. The additional factor $(z_1/z)^n$ corrects for the expansion of the beam with distance from the cathode. According to Eq. (2) $\tau = 0$ implies $\theta = 0$ or $t = (z - z_1)/v_0$. Thus electrons at the center of the bunch move uniformly without acceleration.

The electromagnetic field equations include a set of relations connecting the curl of the magnetic field with the total vector current density (conduction plus displacement). These relations require the divergence of the total vector current density to vanish. In m.k.s. units

$$\frac{\partial}{\partial z} z^n \left[i'(z, t) + \epsilon_0 \frac{\partial}{\partial t} E'(z, t) \right] = 0, \tag{4}$$

or

$$z^n \left[i'(z, t) + \epsilon_0 \frac{\partial}{\partial t} E'(z, t) \right] = J(t). \tag{5}$$

Equation (5) states that the total current is independent of position along the beam.[5] In the absence of an hf return path the total current and also the curl of the magnetic field must vanish. But the curl of the electric field is zero since the field is purely longitudinal. Consequently a magnetic field is not required to com-

[5] F. B. Llewellyn and A. E. Bowen, Bell Sys. Tech. J. 18, 280 (1939).

plete the solution. The electromagnetic field equations possess solutions for systems *a*, *b*, and *c* containing only a longitudinal component of electric field and no magnetic field.

Continuing the analysis under the condition that no hf return path exists we get

$$\epsilon_0 E'(z, t) = -\int_{(z-z_1)/V_0}^{t} i'(z, t')dt', \qquad (6)$$

or

$$\omega\epsilon_0 E(z, \theta) = -\int_0^{\theta} i(z, \theta')d\theta'. \qquad (7)$$

Note that $E(z, 0)=0$ as is required by the uniform motion of the center of the bunch. With the help of Eqs. (2) and (3) the field can be expressed in the form

$$E(z, \theta) = -\frac{i_0}{\epsilon_0\omega}\left(\frac{z_1}{z}\right)^n X(z)\sin\tau. \qquad (8)$$

The value of the electric field prescribed by Eq. (8) must now be inserted into the equation of motion:

$$\ddot{z} = -\frac{ei_0}{\epsilon_0\omega m}\left(\frac{z^1}{z}\right)^n X(z)\sin\tau. \qquad (9)$$

Equation (9) can be integrated formally with respect to t thus expressing z as a function of τ and t. A more convenient form is provided by the substitution $dt=dz/v_0$ which is an exact relation for electrons at the center of a bunch and a good approximation (since $\alpha\ll 1$) for all other electrons. As a result of this substitution the integral form of Eq. (9) begins to resemble Eq. (2). There finally results

$$\theta = \tau - \left[\tfrac{1}{2}\alpha\mu k'(z-z_1) - \frac{i_0 e z_1{}^n}{\epsilon_0 m V_0{}^3} \times \int_{z_1}^{z}(z-z_1)\frac{X(z')}{z'^n}dz'\right]\sin\tau, \qquad (10)$$

in which $k'=\omega/v_0$. The consistency of the calculation now depends on the identity of Eqs. (2) and (10). These equations become identical if the bunching parameter $X(z)$ can be determined as a solution of the integral equation

$$X(z) = \tfrac{1}{2}\alpha k'\mu(z-z_1) - \frac{i_0 e z_1{}^n}{\epsilon_0 m V_0{}^3} \times \int_{z_1}^{z}(z-z')\frac{X(z')}{z'^n}dz'. \qquad (11)$$

Before integrating Eq. (11) it is convenient to introduce a bit of notation. Let

$$h^2 = \frac{i_0 e}{\epsilon_0 m v_0{}^3}, \quad \text{System } a, \quad n=0,$$

$$\Lambda = \frac{i_0 e z_1}{\epsilon_0 m v_0{}^3}, \quad \text{System } b, \quad n=1, \qquad (12)$$

$$\Gamma = \frac{i_0 e z_1{}^2}{\epsilon_0 m v_0{}^3}, \quad \text{System } c, \quad n=2.$$

The parameters h and Λ have the dimensions of a reciprocal length while Γ is a pure number. To obtain numerical values consider the special numerical examples:

System (a)

$i_0=1$ ampere/cm², $V_0=2500$ volts, $h=0.868$/cm;

System (b)

$2\pi i_0 z_1=1$ ampere/cm, $V_0=2500$ volts, $\Lambda=0.120$/cm;

System (c)

$4\pi i_0 z_1{}^2=1$ ampere, $V_0=2500$ volts, $\Gamma=0.060$.

Numerical results for other values of current and voltage are readily found since all three

TABLE I.

System	
Plane (a)	$\frac{1}{2}\mu\alpha k' \sin h(z-z_1)/h$
Cylinder (b)*	$\frac{\pi}{2}\mu\alpha k'(zz_1)^{\frac{1}{2}}[N_1(2(\Lambda z)^{\frac{1}{2}})J_1(2(\Lambda z_1)^{\frac{1}{2}}) - N_1(2(\Lambda z_1)^{\frac{1}{2}})J_1(2(\Lambda z)^{\frac{1}{2}})]$
Sphere (c)	$\mu\alpha k'\left(\frac{zz_1}{1-4\Gamma}\right)^{\frac{1}{2}}\sinh\left\{\frac{1}{2}(1-4\Gamma)^{\frac{1}{2}}\ln\frac{z}{z_1}\right\}$

* The Bessel functions $J_1(x)$ and $N_1(x)$ are tabulated in Jahnke-Emde, pages 157 and 191.

quantities defined in Eq. (12) vary directly as the current and inversely as the three-halves power of the beam voltage.

Equation (11) is best discussed by transforming it into a differential equation. Successive differentiations with respect to z yield

$$\left(\frac{d^2}{dz^2}+h^2\right)X(z)=0, \quad \text{System } a,$$

$$\left(\frac{d^2}{dz^2}+\frac{\Lambda}{z}\right)X(z)=0, \quad \text{System } b,$$

$$\left(\frac{d^2}{dz^2}+\frac{\Gamma}{z^2}\right)X(z)=0, \quad \text{System } c.$$

The boundary conditions are

$$X(z_1)=0,$$

$$\frac{dX}{dz}=\tfrac{1}{2}\mu\alpha k', \quad z=z_1, \quad k'=\omega/v_0. \tag{13}$$

The final results for $X(z)$ are listed in Table I.

Consider now the limiting process $z_1\to\infty$ while $z-z_1$ and i_0 remain fixed. Under these conditions systems b and c are indistinguishable from system a. With the help of the relations $\Lambda=h^2z_1$, $\Gamma=h^2z_1^2$ and the asymptotic values of J_1 and N_1 (Jahnke-Emde, page 138) the second and third lines of Table I are readily reduced to the first line.

The assumption $z\geqq z_1$ has been implicit in the preceding calculation. However it is necessary only to reverse the signs of i_0 and v_0 to obtain the description of systems in which the beam converges toward the axis (b) or the center (c) of symmetry. This change leaves the expressions in Table I invariant in form; they are also negative valued when $z-z_1$ is negative and not too large. It is likely that convergent beams may prove useful in the development of multipliers.

A peculiar property of the spherical beam emerges from the examination of $X(z)$ in Table I. The bunching parameter of a divergent spherical beam is a monatomic increasing function of z when the current is limited by the inequality $4\Gamma<1$ (this requires a total current below 4.16 amperes at $V_0=2500$ volts).

A further point of interest is the total voltage through which the electron falls while traversing the buncher grid space and the drift space:

$$\delta V=\int_{z_1}^{z}E(z',\theta)dz'+\mu V_1\sin\tau, \qquad =\frac{2V_0}{k'}\frac{dX}{dz}\sin\tau. \tag{14}$$

Equation (14) states that the velocity modulation vanishes at points where the bunching parameter attains maximum or minimum values.

It will be observed that the restriction $|X(z)|\leqq 1$ holds for all values of z if α falls below a critical value. Otherwise $X(z)$ may exceed unity for some range $z_2<z<z_3$ and the analysis fails beyond $z=z_2$. In physical terms z_2 is the point at which overtaking or crossing of orbits begins to occur. No progress has been made in evaluating the effect of space charge beyond the initial crossover point.

A simple, but far-reaching, generalization of the preceding analysis may be obtained by the substitution of an arbitrary periodic function $f(\tau)$ for $\sin\tau$. Let

$$f^*(\tau)=\sin\omega t+\sum_2^\infty a_n\sin(n\omega t+\Phi_n) \tag{15}$$

represent the time dependence of the voltage across the buncher grids. Then

$$\mu(s)f(\tau)\equiv\mu(s)\sin\tau +\sum_2^\infty a_n\mu(ns)\sin(n\tau+\Phi_n'). \tag{16}$$

In particular the right-hand member of Eq. (16) may represent the sawtooth function or an approximation to the sawtooth function. If

$$a_n\mu(ns)=\frac{(-1)^{n+1}}{n}\mu(s), \quad \Phi_n'=0 \tag{17}$$

overtaking or crossing of orbits first occurs at $X(z)=2$ coincident with the formation of a perfect bunch (all particles passing the center of the buncher grid space during one cycle pass the fixed point defined by $X(z)=2$ at the same instant). In this case the former restriction $|X(z)|\leqq 1$ should be replaced by $|X(z)|\leqq 2$.

Here and There

Harvard Computation Laboratory

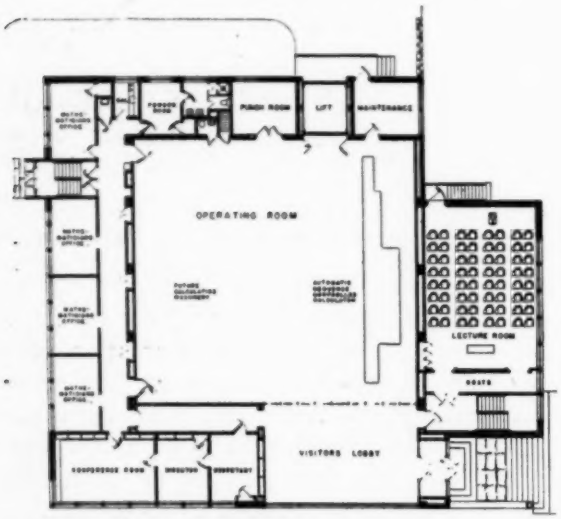
Harvard University's gigantic mechanical brain, the Automatic Calculator, will soon be moved to a new two-story building on the campus. The structure, to be known as the Computation Laboratory of Harvard University, of brick, concrete, and steel, functionally modern in design, is expected to be ready in October.

The Calculator has been doing highly secret war work since it was given to the University by the International Business Machine Corporation in August 1944, and it will continue to operate under a naval contract for two more years. By the terms of this contract, the Calculator now works twenty hours a day, every day in the week, for the Navy, and four hours a day for Harvard. It is under the direction of one of its inventors, Howard H. Aiken, associate professor of applied mathematics.

A constantly whirring device made up of gears, counters, switches, shafts, and control circuits, with 500 miles of wire, 2,000,000 wire connections, and 1464 ten-pole switches in its still-framed body, the Calculator will attain new heights of accomplishment during the next few years. Atomic physics, radio research, optics, electronics, and astronomy are only a few of its fields. With infallible accuracy and great speed, it solves problems which no scientist, however gifted, could live long enough to finish.

The new building will give added scope to these and other activities of the Calculator. Research into the design of instruments for the benefit of science will be furthered through the use of larger quarters, and instruction of graduate students under the new department of Engineering, Sciences and Applied Physics will be increased. Civilian use of the machine will include instruction for students in the design and use of large scale calculators.

In its new home the Calculator will be observed by visitors from a special gallery with a window two stories high. Since it was set up at Harvard, the Calculator has been viewed by more than 2000 visitors, including scientists from Russia, China, France, Brazil, Mexico, Britain, Sweden, South Africa, Cuba, and Holland.



Computation Laboratory for Harvard University, first floor plan.

New Appointments

Several appointments were recently announced by the National Bureau of Standards. Dr. John G. Thompson is now Chief of the Metallurgy Division, and William F. Roeser is Assistant Chief. George A. Ellinger has been designated Chief of the Optical Metallurgy Section, and Harold E. Cleaves is new Chief of the Chemical Metallurgy Section. Dr. Wilmer Souder has been appointed Chief of the Metrology Division, formerly the Division of Weights and Measures, and there are three new section chiefs within this Division: Dr. I. C. Schoonover, Dental Materials Section; Dr. Peter Hidnert, Thermal Expansion Section; and H. Haig Russell, Large-Capacity Scales Section.

Recent Awards

Benjamin Goldberg of New York has been awarded the Exceptional Meritorious Civilian Service Award as a result of his work for the past five years as an optical physicist attached to the technical staff of the Army Engineer Board, Fort Belvoir, Virginia. He won this recognition for his contribution to the technique of metal mirror reflector manufacture and for designing special reflectors for lighting and for other optical devices, particularly for searchlights used extensively by the U. S. Army during the war.

The I.E.S. Medal, given by the Illuminating Engineering Society "in recognition of meritorious achievement which has conspicuously furthered the profession, art, or knowledge of illuminating engineering" has been awarded to Eugene C. Crittenden, Chief of the Electrical Division, National Bureau of Standards, and will be presented to him at the Society's convention in Quebec in September. The most outstanding of Mr. Crittenden's notable achievements in illuminating engineering are his contributions to the development and establishment of acceptable standards and units for the measurement of light.

Army Purchases Atom Smasher

The Manhattan District Project of the Army Engineers, which developed the atomic bomb, has ordered a 100,000,000-volt atom-smashing x-ray machine from the General Electric Company. The machine is being obtained by the Clinton Laboratory at Oak Ridge, Tennessee, which is operated by the Monsanto Chemical Company for the Manhattan District Project.

Auburn Foundation Receives Grant-in-Aid

A Frederick Gardner Cottrell Special Grant-in-Aid has been made to the Auburn Research Foundation of the Albama Polytechnic Institute to support research on the crystal structure of sugars and their derivatives. The grant totals $7500.